

DOCUMENT RESUME

ED 222 755

CE 034 278

TITLE Self-Assessment and Self-Paced Learning Modules for Physician Assistants. The Comprehensive Report.
INSTITUTION American Academy of Physician Assistants, Arlington, VA.; Ohio State Univ., Columbus. Coll. of Education.
SPONS AGENCY Health Resources Administration (DHHS/PHS), Hyattsville, Md. Bureau of Health Professions.
PUB DATE 30 Apr 82
CONTRACT HRA-232-80-0018
NOTE 139p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Allied Health Occupations Education; *Criterion Referenced Tests; Diagnostic Tests; Instructional Materials; *Learning Modules; *Material Development; Pacing; *Physicians Assistants; Professional Continuing Education; Program Effectiveness; *Self Evaluation (Individuals); Test Construction; Test Validity
IDENTIFIERS *Diagnostic Prescriptive Approach

ABSTRACT

A project was undertaken to refine an existing self-assessment system and to develop self-paced learning modules for use in the continuing education of physician assistants (PA). Included among the major activities of the project were the following: (1) modification of the existing criterion-referenced Physician Assistant Self-Assessment Examination (PA-SAE) based on internal consistency and item analyses; (2) refinement of the PA-SAE scoring system; (3) development of a profile of the PA profession based on the results of an examination of a representative national sample of physician assistants; (4) completion of 18 self-paced learning modules based on the results of the PA profile; and (5) design of a continuing competency assurance program. The project has resulted in the production of a viable model for a continuing competency assurance system for physician assistants. To date, 12 learning modules have been completed and 6 more are currently being developed. Recommendations emerging from the project included calls for further consideration of the adequacy of the examination with regard to standards-setting and for development of similar systems for other health professions. (MN)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED222755

SELF-ASSESSMENT AND SELF-PACED LEARNING MODULES FOR
PHYSICIAN ASSISTANTS

The Comprehensive Report

Covers Period of Performance
During
June 24, 1980 - April 30, 1982

Contract No. HRA-232-80-0018

Contractor

American Academy of Physician Assistants
2341 Jefferson Davis Highway, Suite 700
Arlington, Virginia 22202

Contractor's Project Director: Ayres G. D'Costa, Ph.D.
Principal Investigator: C. Emil Fasser, PA-C

Sponsor

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration, Bureau of Health Manpower
Room 4-22, Center Building, Prince George Center
3700 East West Highway, Hyattsville, Maryland 20782

Total Price of Contract: \$127,696

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

✓ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

CE 034278

TABLE OF CONTENTS

	<u>Page</u>
PREFACE	vi
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ABBREVIATIONS AND SYMBOLS	xi
ABSTRACT	1
I. INTRODUCTION	3
1. Purpose of Contract	4
2. Description of Major Tasks	6
II. METHODOLOGY OF TEST DEVELOPMENT AND REFINEMENT	9
1. Initial Test Development Process	12
Domain Specification	12
Scale Definition	14
Item Generation	17
2. Refining the Exam	27
Development of New Items	37
Scoring System Changes	38
Weighted Scoring	38
Competency Standards	39
Error Pattern Scales	42
3. Reliability and Validity	43
Reliability	45
Content Validity	52
Construct Validity	53
Decision Validity	55

4. Exam Reporting System	56
Page One: Letter with Performance Summary and CME Recommendations	57
Page Two: Practice, Need expressed, and Exam Performance Profile	58
Page Three: Actual and Correct Responses to Exam Items	58
Page Four: Role Scales: Raw Scores Rela- tive to Competency Standards	59
Page Five: Role Scales: Quality of Re- sponses	60
Page Six: System Scales: Raw Scores Re- lative to Competency Standards	60
Page Seven: System Scales: Quality of Responses	60
Page Eight: Error Pattern Scales: Raw Scores Relative to Competency Standards .	60
Interpretive Booklet	61
III. NATIONAL PROFILES OF PROFESSIONAL COMPETENCE OF PHYSICIAN ASSISTANTS	70
1. Objectives	70
2. Methodology: Administration to National Sample	72
3. Verification of Representativeness of Sample	75
4. Visual and Verbal Summaries	83
Overall Extent of Competency Attainment . .	84
Specific Strengths and Weaknesses	90

Relationship of Practice Characteristics	
to Competencies Profile	94
Relationship of Perceived Educational	
Needs to Competencies Profile	94
5. Application and Interpretation Guidelines . .	95
Criteria for Learning Module Identification	95
Interpretation Guidelines	100
IV. DEVELOPMENT OF LEARNING MODULES	103
1. Objectives	103
2. Methodology	104
Identification and Prioritization of Areas	
of Deficiency	104
Training Authors	106
Learning Modules Draft: Writing Workshop .	107
Expert Review	108
Developmental Testing	108
3. Results	109
V. DESIGN FOR CONTINUING COMPETENCY ASSURANCE PRO-	
GRAM	111
1. Objectives	111
2. Methodology and Results	112
VI. OVERALL RESULTS AND CONCLUSIONS	117
Recommendations	118
REFERENCES	122

PREFACE

The Contractor for this project was the American Academy of Physician Assistants, a national association representing 13,000 Physician Assistants in the U.S.A. and with its national offices at 2341 Jefferson Davis Highway, Suite 700, Arlington, Virginia 22202. The following persons are identified as the Contractor's staff involved in the performance of this project:

C. Emil Fasser, PA-C, Principal Investigator
Peter Rosenstein, Executive Director
Robert Johnston, Comptroller
Gale Emery, Liaison Associate

The performance of this entire project was sub-contracted to The Ohio State University, College of Education, Department of Educational Foundations and Research. Project offices were located at 1250 Chambers Road, Suite 250, Columbus, Ohio 43212.

The following persons are identified as the Sub-Contractor's staff involved in the performance of this project:

Ayres G. D'Costa, Ph.D., Project Director
John Belland, Ph.D., Project Consultant for
Modules Development
Lenore Hinman, M.A., Administrative Associate,
1980-81
Michael Lower, M.S., Research Associate, 1980-81
Wendy McColskey, M.S., Research Associate,
1981-82
Joan Watson, M.S., Research Associate, 1980-82
John Tipka, Fiscal Manager
J. Thomas Targoss, M.S., D.D.S., Illustrator
Elizabeth Rhyner, Secretary
Terri Spauling, Word processing

The following persons served as consultants and on the major Committees of the project.

Advisory Committee

J. Rhodes Haverly, M.D. - National Commission on
Certification of Physician's Assistants
Joseph A. Intile, M.D. - American Society of
Internal Medicine
Calvin F. Kay, M.D. - American College of Physi-
cians
Robert G. Pierleoni, Ed.D. - Association of Amer-
ican Medical Colleges
Robert H. Shackelford, M.D. - American Academy of
Family Physicians
Richard E. Wilkes, PA-C, M.S. - Association of
Physician Assistant Programs
Judith Willis, PA-C, M.A. - American Academy of
Physician Assistants
Charles A. Worley, M.D. - American Academy of
Family Physicians

Learning Module Authors

A. Sherwood Baker, M.D. - Family and Community
Medicine, University of Missouri
Raymond Brock, PA-C, M.Div. - V.A. Medical
Center, Hampton, Virginia
C. Emil Fasser, PA-C - Baylor College of
Medicine, Houston, Texas
Geoffrey R. Grambau, M.D. - Department of Pul-
monary Medicine, Michigan State University
J. Jeffrey Heinrich, PA-C - Yale University
School of Medicine, New Haven, Connecticut
Lawrence Kammer, PA-C, M.S. - Educational Coor-
dinator at Mercy College, Physician Assistant
Program, Detroit, Michigan
Ann Teske, R.N., Ph.D. - Chief of Nursing Ser-
vice, Health One, Columbus, Ohio
Wendell W. Weber, Ph.D., M.D. - University of
Michigan, Ann Arbor, Michigan
Judith Willis, PA-C, M.A. - Southwestern Michi-
gan Health Education Center, Kalamazoo, Mich-
igan

LIST OF TABLES

	<u>Page</u>
1. Major Responsibilities within Areas of Competence	15
2. The Seventeen Role Scales	19
3. The Twenty-nine Body System Scales	21
4. Recommended Critical Incidents in Each Body System Scale	28
5. Reliability of System Scales: Weighted Versus Dichotomous Scoring	40
6. The Thirteen Error Pattern Scales	44
7. Role Scales: Reliability Indices	47
8. System Scales: Reliability Indices	48
9. System Scales: Validity Indices	49
Comparison of SAE-PAP and APAP-NS Responses	
10. Breakdown of PAs by Sex	78
11. Year of Graduation	78
12. Medical Background Prior to Training	79
13. Location of Practice	79
14. Size of Community where Practice is Located	80
15. Type of Practice	80
16. Type of Patient Care	81
17. In-Hospital Location for Patient Care	81
18. Entry Level Standard and Reliability Index for Role Scales	85

19. Entry Level Standard and Reliability Index for System Scales	86
20. Entry Level Standard for Error Pattern Scales .	87
21. List of Practice Characteristics in Rank Order .	96
22. List of Perceived CME Needs in Rank Order . . .	97
23. Module Data Analysis	98
	& 99

LIST OF FIGURES

	<u>Page</u>
1. The 3D Model of PA Competence	18
2. Map of the 17 Role Areas X Body System Scales .	20
3. Map of the 28 System Scales X Medical Inter- vention Scales	22
4. Guide for Generating Item Scenarios	25
5. Sample Scenario	26
6. Individual Report	62
7. Number of Role Scales Passed at Competency Level by SAE-PAP Sample	88
8. Number of System Scales Passed at Competency Level by SAE-PAP Sample	89
9. Percent of PAs who 'Passed' Each Role Scale at Entry-Level Competency	91
10. Percent of PAs who 'Passed' Each System Scale at Entry-Level Competency	92
11. Percent of PAs who 'Exceeded' Advanced-Level Error Pattern Standard	93

LIST OF ABBREVIATIONS AND SYMBOLS

AAPA	American Academy of Physician Assistants
PA	Physician Assistant
CME	Continuing Medical Education
NCCPA	National Commission for the Certification of Physician Assistants
HHS	Health and Human Services
HRA	Health Resources Administration
OSU	Ohio State University
PA-SAE	Physician Assistant Self-Assessment Exam
E	Exam Scores
P	Practice Profile
N	Expressed Need
APAP-NS	Association of Physician Assistant Programs- National Survey
SAE-PAP	1981 AAPA Self-Assessment Examination Profile
NS	Not Significant
CA	Coefficient of Acceptance
SEM	Standard Error of Measurement

ABSTRACT

Continuing medical education has been operationalized by the American Academy of Physician Assistants as a self-assessment system. Improving on this system was the central purpose of this contract (HRA-232-80-0018). It involved refining a criterion-referenced Physician Assistant Self-Assessment Exam (PA-SAE), profiling the PA profession using this exam, developing self-paced learning modules for PAs, and providing a design for the continuing competency assurance system.

The Role Delineation undertaken under previous contracts (HRA-231-75-0209; HRA-231-76-0053) resulted in the identification of 11 major areas of physician assistant responsibility. These areas were taken as a starting point in attacking the task of classifying the domain to be covered by the PA-SAE. The end result of this task was that the PA-SAE should contain two distinct sets of reporting scales (17 Role Scales, and 28 Body System Scales).

Refining the PA-SAE consisted of revisions and additions to the scales mentioned above. Internal consistency and item analyses were run. From these data, decisions were made to delete, revise, or write items. This refined set of items represented Form A4 of the PA-SAE.

Three other refinements were added to the computerized

reporting system for Form A4. These were error pattern scales, a weighted scoring routine, and a two-tiered cut-off system (one set for entry level PAs and one for experienced PAs).

The criterion-referenced exam was administered to a representative sample of PAs. The responses were summarized to provide a continuing education need profile which could be broken down by such variables as geographic location and practice setting.

The topics covered by the Self-Paced Learning Modules were derived by critically examining the results from the administration of the PA-SAE to a sample of PAs, representing a profile of the profession. These modules were designed to allow PAs to work on improving their knowledge in areas essential to their practice.

The PA-SAE, the national PA competency profile, and the learning modules are all part of a life-long competency assurance program for PAs. All three of these components are part of a process begun with the professional needs of PAs in mind. A description of how these components fit into a larger competency assurance system involving cyclic recertification was the fourth component of this project.

I. INTRODUCTION

National health policy has shifted in recent years from an emphasis on health manpower or services development to a concern for practitioner competence in the delivery of health care. Conant (1978) pointed out that the Health Training Improvement Act of 1970 (P.L. 91-519) provides the basis for the major steps taken by the Division of Associated Health Professions (Health Resources Administration, Department of Health and Human Services) toward assuring the competency of allied health practitioners. These steps have been reflected in a series of projects supported by the Division and designed to result in a program of credentialing allied health professionals so as to assure the public of the continuing competence of these practitioners in providing quality health services. Major phases of this program of credentialing to date have been role delineation and verification, the identification of professional skills and curricular objectives, the development of criterion-referenced self-assessment examinations, and the administration of these examinations by the respective health professions for the identification of their competency strengths and weaknesses. The present project, involving the refinement of a criterion-referenced self-assessment examination and the development of a prescriptive and self-paced learning system

based upon a national profile of the profession, was considered a legitimate and significant next step in the continuation of this national competency assurance program by the Division.

Purpose of the Contract

The purpose of this contract was to extend the thrust of the continuing competence assurance program begun in the 1970's, when the American Academy of Physician Assistants (AAPA) adopted a policy of certification, mandatory continuing education, and recertification every six years as one means of assuring the competency of physician assistants (PAs) in their professional performance. To maintain a valid certificate, the PA must provide evidence of 100 hours of approved continuing medical education (CME) every two years. The accreditation of CME programs and the logging of CME credit earned by each PA is performed by the AAPA. Previous programmatic efforts by the AAPA have developed a delineation of the PA Role (Contract HRA 231-75-0209), and more recently a verification of the role delineation and the development of a Self-Assessment Examination to facilitate the continuing competency of PAs (Contract 231-76-0053).

The Self-Assessment Examination (SAE) developed under the previous project was conceived by the AAPA as part of an organized system of CME to be made available to its members. The first step of this system, the criterion-refer-

enced exam, was developed to provide diagnostic feedback to the PA regarding strengths and weaknesses in areas relevant to clinical practice. In other words, the exam was designed to inform the PA as to where he or she stands relative to the domain of content considered essential for the competence of PA practitioners in the field. Learning prescriptions, in the form of references to texts or articles, were planned for inclusion as part of the exam feedback. One purpose of the present project was to refine this self-assessment exam (SAE) developed under Contract HRA 231-76-0053, so that the examination and its prescriptive feedback would have a sound basis.

A second purpose of the project was to extend the prescriptive aspect of the self-assessment system to include references to a set of specially-developed self-paced learning modules. An important step in the continuing education process is prescribing sources of remediation for PAs who have been identified by the self-assessment exam as deficient in an area(S). Towards this goal, this project was contracted to develop Self-Paced Learning Modules that would be made available to PAs upon request subsequent to receiving feedback about their performance on the exam.

In order to identify content areas of relative critical need to be covered by the learning modules, a national profile of the profession was undertaken as the third purpose of the present project. After administering the SAE to a representative sample of PAs, a profile of the profession

was described by summarizing the sample's performance on the areas (scales) included in the exam. Areas in which the group were deficient was considered as prime topic areas for which modules should be developed.

The self-assessment exam, the national PA competency profile, and the development of needed learning modules are three parts of a life-long competency assurance program for PAs. These components are part of a systematic CME process begun with the professional needs of PAs in mind. The description of how these components fit into a larger competency assurance system involving cyclic recertification was identified as the fourth purpose of this project.

Description of Major Tasks

As described in the above section, the project set out to accomplish four major tasks:

1. Refine the Self-Assessment Instrument
2. Profile the Physician Assistant Profession
3. Develop Self-Paced Learning Modules for Practicing Physician Assistants
4. Design a Continuing Competency Assurance Program

1. Refine the Self-Assessment Instrument

Refinement was operationalized as the production of a new and improved form (A4) of the self-assessment exam. Thus, the emphasis was on revising and adding items, and on

experimenting with improvements to the scoring system. One addition to the scoring system was error pattern scales, such that each time an examinee made a wrong response, the wrong response was classified as a certain kind of error. Scores were then generated for the frequency of these different types of errors over the whole test. A second addition was a weighted scoring system, which supplemented the dichotomous scoring system. A third addition was a two-tiered system of cut-offs or standards for each scale, such that one standard is relevant to entry-level PAs and the other is relevant to experienced PAs. The methodology for all these additions involved obtaining judgements from a committee of experts, and subsequently, changing the computerized scoring system.

2. Profile of the Physician Assistant Profession

This task involved generating a representative sample of PAs from the AAPA membership list, requesting these PAs to take the SAE, and analyzing the responses to identify the strengths and weaknesses of the profession. In particular, the frequency of PAs passing and failing the 17 PA Role scales and the 29 Body System scales was of interest. The frequencies of sub-groups of PAs, represented by such variables as geographic location and practice setting, were also obtained to verify the representativeness of the sample.

3. Develop Self-Paced Learning Modules for Practicing Physician Assistants

The first component of this task was the decision as

to what topic or content areas should be covered by the modules. This decision was made on the basis of the data gathered from the administration of the SAE to the representative sample of PAs. In particular, scales on which large percentages of PAs failed the exam were noted. Once topic areas were identified, authors were obtained, and training and writing workshops were held. As the modules were written they were developmentally tested, reviewed by experts, and edited.

4. Design a Continuing Competency Assurance Program

This task focused on producing a description of how the self-assessment system fits into the larger system of certification and recertification.

II. METHODOLOGY OF TEST DEVELOPMENT AND REFINEMENT

Tests are used for different purposes, and the purpose has implications for the path followed in test development. The goal of the SAE was to provide the most informative feedback as possible regarding the PA's strengths and weaknesses. That is, the test was designed to let the PA know where he or she stands relative to the domain of content (knowledge or skills) considered essential for PA practitioners in the field. Informing the examinees about how they did relative to others who took the test (norm-referenced information) is not as important as informing them how they stand relative to an agreed upon national standard for PA competence (criterion-referenced information).

These two different types of score interpretation, (norm-referenced and criterion-referenced), lead to two different emphases in test development. A criterion-referenced test classifies students as masters or non-masters of a specified content or skill domain. Thus, specifying the domain accurately and developing items that are congruent with this domain are of primary importance. A norm-referenced test is primarily concerned with the ordering of examinees, or showing where they stand relative to others taking the test. The primary concern in developing this kind of test is choosing items that discriminate between people. The domain in this kind of test may not be as clearly specified.

The criterion-referenced interpretation, and therefore, adequate domain specification, is of primary concern in the development of the SAE. In addition, because the purpose of the test is diagnostic, it was thought that the more information the PA received the better. For this reason, both criterion and norm-referenced interpretations are provided by the test. The PA receives feedback about his/her performance on each of 17 Role scales and 29 Body System scales relative to the absolute standard set for each scale. The PA also receives a normative profile of his/her patient load, expressed needs, and exam scores. In this case, each index is presented in a stanine form, which describes where the examinee stands relative to peer performance.

The test development process was adapted from that suggested for criterion-referenced tests (Hambleton and Eignor, 1979). Typically, objectives stating what the test should measure (domain specifications) are prepared first. Once the broad content of the test is outlined, test specifications are identified. Additionally, the item formats, and the instructions to item writers must be clarified. Then the test items are written so as to measure the objectives or specified domain. At this point, consideration should be given to the match of the items to the content that they are supposed to measure, and to their representativeness. The items are edited and the test is assembled. Next is the important step of setting the standards used to interpret examinee performance. The test is administered.

Data concerning reliability and validity are collected and analyzed, and Test Users and Technical manuals are prepared.

The Self-Assessment Exam (SAE) was initially developed under Contract HRA-231-76-0053. The large job of domain definition was undertaken under this previous project. Test specifications were clarified, items were written and matched to appropriate scales, and standards were set.

The present project began at the item refinement stage. That is, items written as part of the test produced under the previous project were evaluated, and the final form A4 of the test became available. This section of the Report begins with a quick overview of the initial test development process of the previous project, explaining how the test domain and the reporting scales were identified, and the rationale behind the critical incident technique of item generation. Following that discussion is a description of the refinement of the exam, which is one of the products of the present project and covers topics such as development of new items, scoring system changes, weighted scoring, competency standards, error pattern scales, and test interpretation. The technical quality of the new exam is then discussed under a section titled: Reliability and Validity. Several concepts of reliability and validity were pursued, with special emphasis on concepts appropriate to criterion-referenced tests relevant to the self-assessment instrument.

1. Initial Test Development Process

Domain Specification

Popham (1978) states that criterion-referenced tests are designed to ascertain an individual's status with respect to a well-defined domain. The precise definition of the domain in terms of skills is critical to the concept of criterion-referenced testing' because of the need to make generalizations about the mastery or non-mastery of these skills based on test scores. Domain definition is important for the interpretation of test scores, but for this exam to serve the diagnostic purpose, it was also important to define the domain in such a way that the linkages between the test results and the PA's experience in the field is accurate. Only then will a PA recognize the implications of his or her weakness in some skills or knowledge as identified by the SAE in relation to his or her ongoing practice. This consideration is an important one in self-assessment procedures. To be relevant, the domain definition should reflect the skills and knowledge used by PAs.

The domain of this self-assessment exam is the performance expected of a competent entry-level PA. Wilson (1976) argues that competency assurance in a credentialing program must be based on a role delineation. In other words, assessment must be based on a sound classification and description of what PAs actually do. Fortunately, such an

analysis (role delineation) has been previously completed in the case of the PA profession by the AAPA.

The end result of a role delineation of a profession is a detailed list of the tasks which a practitioner is expected to be competent to perform. This contrasts with the other approaches, such as task analysis, which list all tasks that a practitioner actually performs. Role delineation facilitates a position classification and enables a minimum standard to be set of skills expected of all practitioners in the profession. Furthermore, a role delineation is expressed in terms of performance responsibilities, rather than just knowledge expected, thereby ensuring job-relatedness.

The methodology of the role delineation began with generating a list describing the tasks PAs perform. A sample of recent PA graduates was selected and asked to check the tasks on this list they typically perform as practitioners. Percentages of this total sample that checked each task were calculated. These percentages of PAs who typically perform each task were used by an expert Committee, along with the task's criticality to health care, as criteria to decide whether or not a task belonged in the verified role delineation. A revised checklist was sent to the same sample who completed the first checklist and to a new sample who had been in practice several years. The Committee again used these responses, along with their expert opinion to establish a final list of responsibilities appropriate for and

to be expected of a practicing PA.

The role delineation contains three comprehensive types of roles or areas of competence: the professional, interpersonal, and the clinical. Within the role delineation, major and specific responsibilities describe in progressive detail the particular tasks a practitioner should be able to perform competently. The eleven major responsibilities describe broad areas of performance, each conveying one major activity of the role. The major responsibilities are broken down into lists of more specific responsibilities. The eleven major responsibilities within the three areas of competence are reproduced in Table 1. These responsibilities represent relevant and critical competencies expected of a PA. This role delineation then was a natural starting point when work began on defining a domain for a criterion-referenced self-assessment tool.

Scale Definition

The starting point for classifying the domain for the self-assessment exam was the role delineation study. Upon consideration it was decided that the 11 role areas, although critical to PA competence, might not be easily translated into test items nor appropriate for in depth feedback. That is, concern was expressed over the difficulty of measuring attitudinal and interpersonal skills, such as demonstrate professional behavior, promote acceptance of PA role, by paper and pencil methods. Another consideration was that

TABLE 1
MAJOR RESPONSIBILITIES WITHIN AREAS OF COMPETENCE



the broad role areas may not be specific enough to serve as the best feedback mechanism for continuing education purposes. For example, being told that one is deficient in data-gathering skills may be too global to serve as a meaningful remedial prescription for a PA.

It was the judgement of the Committee that Body Systems represented the most useful criterion to use in scale development for several reasons: 1) the study of medicine is based on understanding a disease entity whose effects may be classified by body systems; 2) Body systems provide a better reference with respect to medical care interventions, such as, emergency, acute, chronic; 3) it is easier to understand the major and specific responsibilities of the PA's clinical role based upon a body system orientation; and 4) such a classification process allows for analysis of and feedback on PA performance regardless of practice type.

Three major dimensions or categorizations were identified as relevant to the identification of scale systems for feedback purposes. These were PA Role, Body Systems, and Medical Intervention. Two sets of scales were derived from these three dimensions: Role Scales which combined PA Role with Body Systems, and Body System scales which combined Body Systems with Medical Interventions.

The 13 Body System areas selected include endocrine, eyes & otorhinolaryngology, respiratory, cardio-vascular, hematology, gastro-intestinal, genito-urinary, reproductive, neurology, psychosocial, and an other category including

economic and diet issues. The four Medical Interventions are Emergency, Acute, Maintenance, and Chronic. See Figure 1.

The end result of the scale definition effort was two distinct sets of reporting scales. Taking the 11 major role areas as a starting point, the following responsibilities, within the clinical role, were recognized as important: gathering data, analyzing data, formulating plans, and managing patients. These responsibilities when matched with the 13 body system areas listed above expanded the number of Role Scales to 17 (Table 2 and Figure 2). These 17 Role Scales were one set of reporting scales.

A second set of reporting scales was developed to represent a content or knowledge classification system rather than a process or role classification system. The Committee felt that the nature of the presenting illness and thus, the type of medical intervention necessary, were also important aspects of PA performance to be considered when assessing competency. The four identified Medical Intervention types were then crossed with the 13 Body System areas to produce the 29 Body System Scales. See Table 3 and Figure 3. The selection of items for these two sets of scales is discussed in the next section.

Item Generation

The traditional approach to item generation in criterion-referenced tests involves the development of a Test

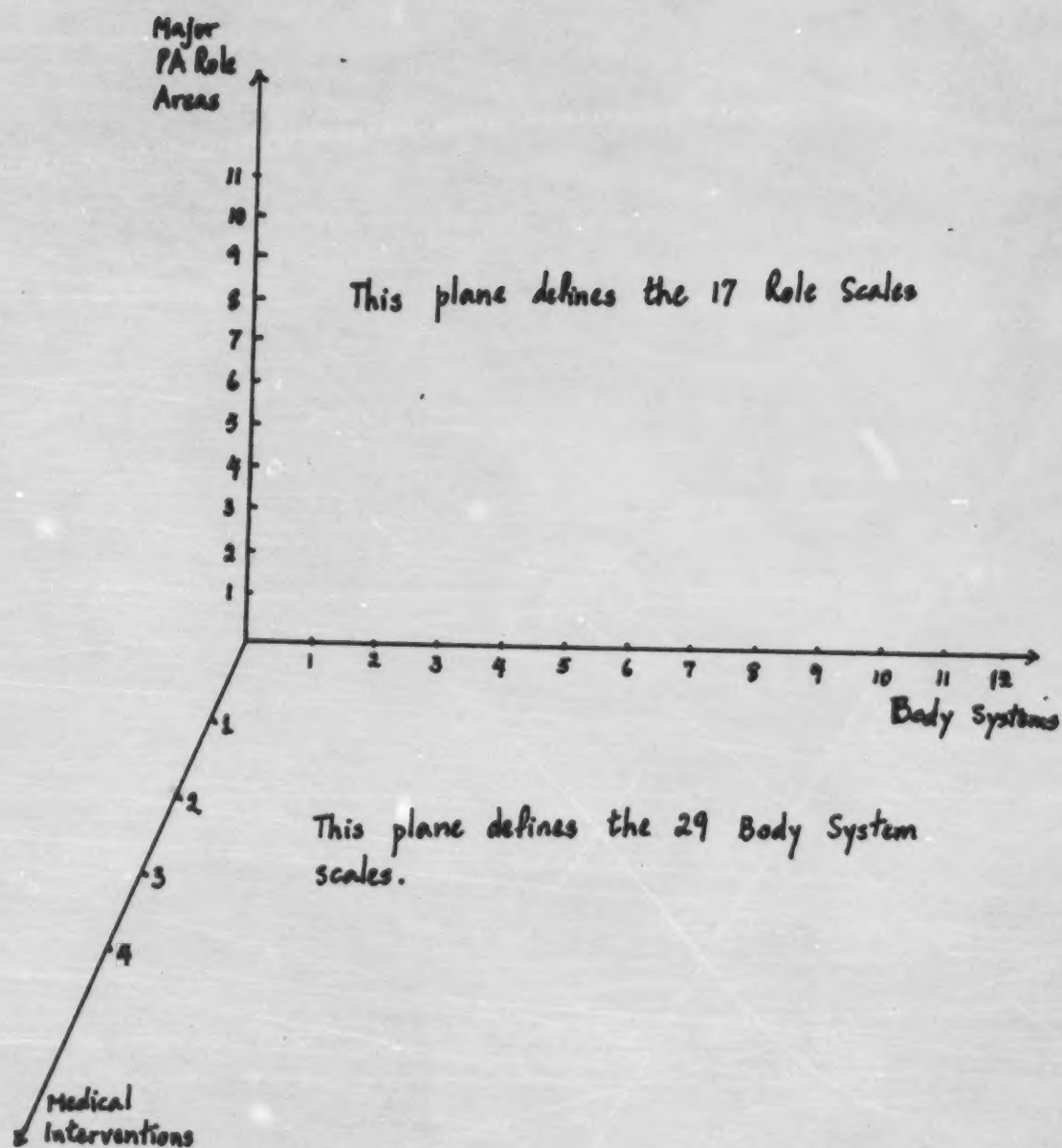


FIGURE 1. A 3-D MODEL OF PA COMPETENCE

TABLE 2
THE 17 ROLE SCALES

1. Professional Role
2. Interpersonal Behavior
3. Gather Data-Respiratory, Cardio-vascular
4. Gather Data-Gastro-intestinal, Genito-urinary
5. Gather Data-Psycho-social, Neurology
6. Analyze Data-Endocrine
7. Analyze Data-Respiratory, Cardio-vascular
8. Analyze Data-Hematology
9. Analyze Data-Gastro-intestinal, Genito-urinary
10. Analyze Data-Psycho-social, Neurology
11. Formulate Plan-Musculo-skeletal
12. Handle Dermatology Problems
13. Manage Patients-Respiratory, Cardio-vascular
14. Manage Patients-Gastro-intestinal, Genito-urinary
15. Manage Patients-Reproductive
16. Manage Patients-Psycho-social, Neurology
17. Handle Pharmacology Problems

FIGURE 2

MAP OF THE 17 ROLE AREAS X BODY SYSTEM SCALES

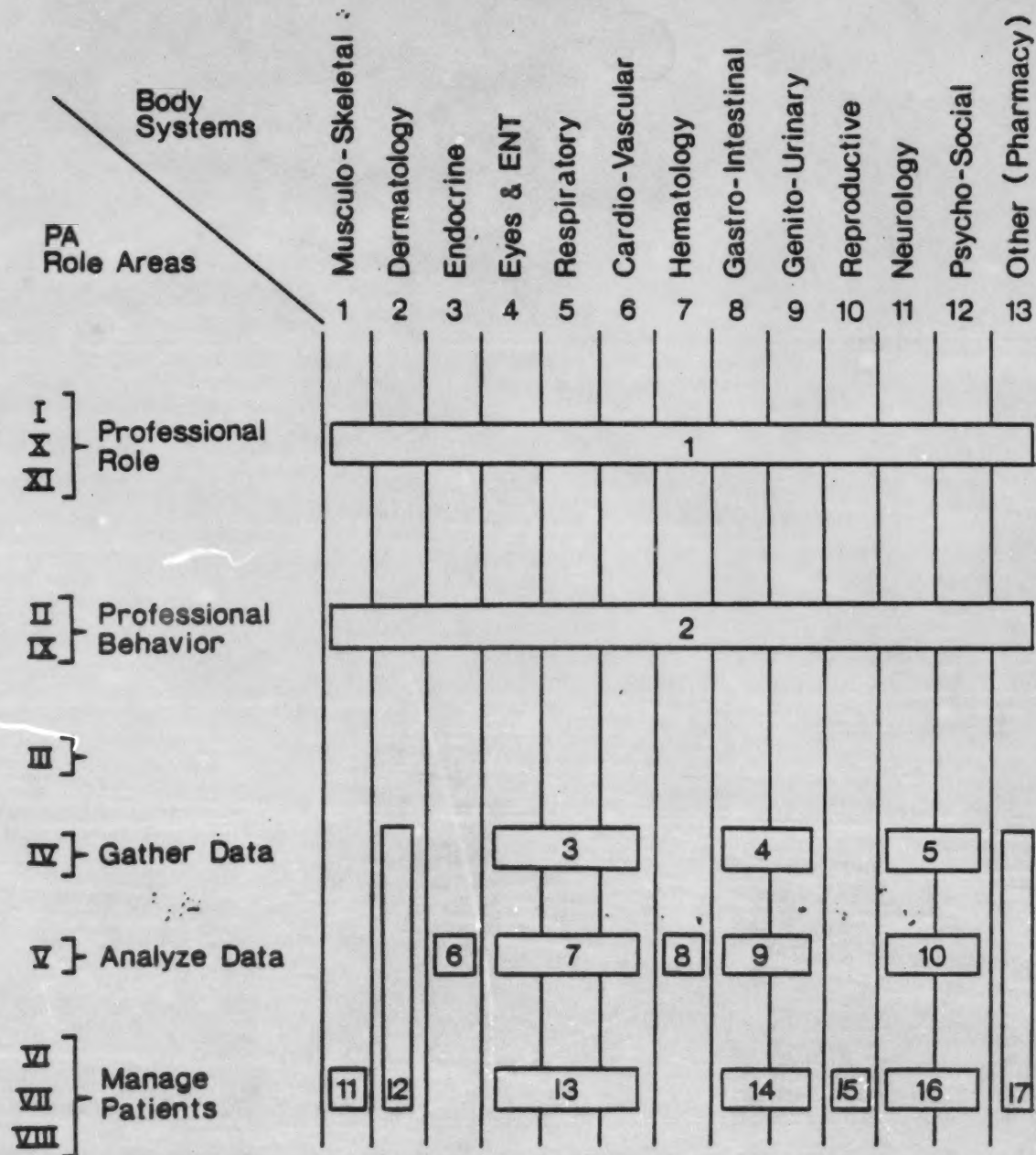


TABLE 3

THE 29 BODY SYSTEM SCALES

EMERGENCY

1. Cardiovascular
2. Respiratory
3. Gastrointestinal
4. Neurology

ACUTE

5. Cardiovascular
6. Musculoskeletal
7. Respiratory
8. Gastrointestinal
9. Genitourinary
10. Neurology
11. Endocrine
12. Dermatology
13. Eyes & ENT
14. Hematology
15. Reproductive
16. Psychosocial

CHRONIC

17. Cardiovascular
18. Musculoskeletal
19. Respiratory
20. Gastrointestinal
21. Genitourinary
22. Neurology
23. Endocrine
24. Dermatology
25. Eyes & ENT
26. Hematology
27. Reproductive
28. Psychosocial
29. Other (Econ, Diet)

FIGURE 3

MAP OF THE 28 BODY SYSTEM x MEDICAL INTERVENTION SCALES

	Emergency	Acute & Maintenance	Chronic
1. Musculo-Skeletal		6	18
2. Dermatology		12	24
3. Endocrine			23
4. Eyes & ENT		13	25
5. Respiratory	2	7	19
6. Cardio-Vascular	1	5	17
7. Hematology		14	26
8. Gastro-Intestinal	3	8	20
9. Genito-Urinary		9	21
10. Reproductive		15	27
11. Neurology	4	10	22
12. Psycho-Social		16	28
13. Other (Pharmacy)		11	

Specification Matrix, usually in a "content areas" by "cognitive objectives" format. The type of item in terms of content and format is then specified in advance of item writing by a set of item generation rules (i.e., "amplified objectives," Popham, 1978). The purpose of this pre-specified type of item writing is to guarantee that the area of competence being measured is accurately translated into test items. If the items are accurate representations of the skill or competence being measured, then accurate interpretations of a score on these items can be made (Popham, 1976). That is, explicit test specifications and congruent test items lead to an accurate interpretation of a criterion-referenced test score.

The major item generation approach utilized for the SAE was derived from the critical incident technique first proposed by Flanagan (1954). This approach was favored over other item generation plans because the domain of professional competence is different from "scholastic achievement skill"-type objectives. Professional competence defined in terms of critical incidents relevant to role responsibilities was thought to be a more relevant starting point than the usual representation by performance task objectives.

A critical incident technique is defined by Flanagan as an extreme behavior, either outstandingly effective or ineffective with respect to attaining the targeted goal (such as competence at a given task, skill, or role). The job of identifying critical incidents fell on the item

writers. Thus, the item writers were given more responsibility using this critical incident approach than with an amplified objectives approach per se. The item writers were given raw data from the role delineation study of PAs, and were then asked to mentally analyze each role responsibility and recall incidents within it that are critical and could make the difference between a competent or incompetent PA. Figure 4 presents the instructions given to the item writers for the purpose of generating critical incidents for each role responsibility. Figure 5 presents a sample Scenario which illustrates the format into which the critical incident generated is organized by the item writer. The critical incident description includes a list of the major skills needed, the most likely errors that might occur in handling the incident, and the education prescriptions recommended for remediation.

The relevant outcomes from this item generation process include: 1) a Scenarios file which is a compilation of critical incidents in the format shown in Figure 5, and 2) a test item bank which provides a record of the psychometric development of each item. The importance of the Scenarios file is that many items can be generated from each scenario by altering the conditions (such as patient age or sex) and the types of errors represented in the incorrect options. The present project, in refining the SAE, augmented the Scenarios file and the test item bank so as to improve the relevance of the SAE to PA competence assessment.

FIGURE 4

GUIDE FOR GENERATING TEST ITEM SCENARIOS¹

I. Study the Role Delineation Model to Identify/Think of Critical Incidents

Note: A critical incident is defined as a set of behaviors that characterize either effective or ineffective performance. Identifying these extremes of a performance dimension in terms of critical incidents helps to understand and to define the performance dimension for measurement purposes.

- 1.1 Pick an item from the Model. Start with #1A. (Role Area 1, Sub-area A). Accept that the role of the PA is limited by supervising physician, legal limitations, etc.
- 1.2 Think of a situation/incident in which a PA very effectively accepted his/her role limitations

OR

Think of a situation/incident in which a PA handled his/her role limitations very ineffectively.

II. Write a Test-Item Scenario for this PA Critical Incident

- 2.1 Describe generally what happened during the incident.
- 2.2 Define the conditions in which this incident unfolds:
 - 2.21 The location/setting (hospital, office, etc.)
 - 2.22 The other health professional involved
 - 2.23 The type of patient involved (sex, age, socioeconomic status, disposition, clinical condition, etc.)
 - 2.24 The type of health care situation involved (preventive, remedial, rehabilitative, etc.)
- 2.3 List the major skills that the PA needs to handle this situation effectively.
- 2.4 List some typical errors, mis-cues, slip-ups that a PA might succumb to in this situation.
- 2.5 List some remedial learning prescriptions that you would recommend in the case of each error.

¹Prepared by Dr. Ayres D'Costa, College of Education, The Ohio State University.

Figure 5

Critical Incident Description

A 4 year old female child is accompanied by her 30 year old mother. This is the oldest of two children in an average, middle class family. The mother is concerned about the child's speech indicating that neighbors and friends as well as the family often cannot understand her. The mother believes that not being understood will emotionally upset the child and lead to developmental problems.

Physically the child is a normally developed 4 year old. Further history reveals the child began to attempt speech at about nine months however, clarity was always a problem. The child seems reluctant to initiate conversation. She will answer questions, but is more difficult to understand than the average 4 year old.

Conditions:

System category - Language development
Patient age - pediatric
Patient sex - female

Skills Needed:

Language and social history
Physical exam
Growth and development assessment

Errors Most Likely:

Incomplete language, social, and physical exam
Inaccurate growth and development assessment
Inappropriate referral
Misdiagnosis and inappropriate treatment

Hambleton and Eignor (1979) suggest that an important area of concern in validating the choice of items for a criterion-referenced test is whether or not the items adequately sample the particular domain. With the critical incident approach, the question is answered both in terms of criticality and frequency (D'Costa and Watson, 1981). Criticality implies that the identified behavior makes a difference in terms of the performer being judged as competent or incompetent. The use of critical incidents rather than representative incidents should add parsimony and more powerful validity to the exam. Frequency implies that the critical incidents should represent critical and frequently occurring incidents rather than esoteric, non-frequently occurring incidents. These are judgements that were made first by the item writers who chose critical incidents and secondly by an expert Advisory Committee who reviewed the incidents and the items. Table 4 presents a list of the critical incidents considered essential for the content domain of the PA Self-Assessment Exam.

2. Refining the exam

One of the four major tasks of the present project was that of refining the self-assessment exam. Refining the exam was attacked on several fronts, with special focus on item quality and on the precision/effectiveness of the scoring system.

TABLE 4

RECOMMENDED CRITICAL INCIDENTS IN EACH BODY SYSTEM SCALE

Scale1

Cardiovascular Emergency

Myocardial Infarction symptoms

4 Coronary Insufficiency

Pulmonary Edema treatment

1-2 PVC's treatment - drug & dose

Vent. tachycardia recognition

Vent. tachycardia treatment

Arrhythmias (Pacer & EKG)

Syncope

Stroke

2

Respiratory

Injury - establish airway

2 Status asthmaticus therapy and complications
treatment pneumothorax

1 Pneumothorax symptoms and recognition

Pulmonary embolism

Conditions warranting emergency tracheostomy

1 Infection

Emphysema

Apnea in newborn

3

Gastrointestinal - Emergency

trauma treatment

2 GI hemorrhage

2 Appendicitis - diagnosis and treatment

2 Diagnosis abdominal pain

Indications for gastric lavage

Duodenal ulcer

Hematemesis

Pediatric vomiting, diarrhea, dehydration
management4

Neurology - Emergency

Status epilepticus

Motorcycle accident - discharge instructions

Relationship between head and cervical spine
injuries

Orbital fracture

Minor auto accident - confusion - diagnosis

Occurrence of febrile convulsions

Coma

Symptoms Meningitis

Hemianopsia
 Stroke/paralysis
 Aphasia
 Spinal cord trauma

- 5 Cardiovascular Acute & Maintenance
 Diagnosis rapid heart rate
 Infant 1 hour old - diagnosis
 Lab work for 12 year old with systolic murmur
 Peak action time for Nitroglycerine
 Action for digitalis toxicity with CHF
 Identification of illustrated murmur
 Recognition of CHF symptoms
 2 Treatment of hypertension
 Diagnosis cerebral thrombosis
 Palpitation and agitation - data
 2 Rheumatic fever
 Explanation of murmur
 Angina
 Arrhythmia
 TIA
- 6 Musculoskeletal - Acute & Maintenance
 laceration over left eye
 2 treatment of fall on outstretched hands
 2 identification and treatment of right hip
 contusion
 2 treatment shoulder pain
 fracture arm 17 month old
 gonococcal arthritis
 auto accident treatment possible cervical
 spine injury
 football knee injury initial treatment
 identification of injury knife laceration to
 hand
 Colles' fracture treatment for 7 year old
 low back pain identification
 Shoulder injury
 Muscle spasm
 General fractures
 Tendonitis
 Dislocation
 Sprain/strain
 Arthritis
- 7 Respiratory - Acute & Maintenance
 identification dyspnea following (4 weeks) MI
 2 causes of pediatric asthmatic symptoms
 asthmatic attack prevention
 asthmatic attack symptoms
 Respiratory distress analysis of ABG
 Influenza symptoms
 Symptoms of pleural effusion
 TB skin test conversion - treatment

Medicines for acute asthmatic attack
 Symptoms post operative atelectasis
 lab test for acute respiratory problem
 admitted to hospital
 Bronchitis/emphysema
 COPD
 Allergy
 Pneumonia
 URI

- 8 Gastrointestinal - Acute & Maintenance
 Data collection - hepatitis
 Lab studies suspected hepatitis
 Management hepatitis
 Causes of diarrhea in 4 year old
 Diagnosis appendicitis
 Colitis as toxic reaction
 Symptoms pancreatitis
 treatment pancreatitis
 Diagnosis 22 year old anorexia, fatigue and
 sore throat
 Management 22 year old anorexia, fatigue and
 sore throat
 Diagnosis dehydration
 Referral for diverticulitis
 treatment of pain in reflux esophagitis
 symptoms of pain in reflux esophagitis
 occurrence of pain in duodenal ulcer
 data collection in diarrhea
 cause black stools in infant
- 2 Data analysis epigastric pain
 Diagnosis projectile vomiting in 1 month old
 Diagnosis CA colon
 Diagnosis choledocholithiasis
 Vomiting/dehydration
 Management & data
 Cholecystitis
 Influenza
 Peptic disease
 Colitis/diverticulosis
- 9 Genitourinary - Acute & Maintenance
 Cause dark brown urine
 Cause of gross hematuria 25 year old male
 Prostatic hypertrophy recognition
- 2 Prostatic hypertrophy treatment
 Data collection urinary frequency
 Diagnosis urinary frequency
 Diagnosis vaginal discharge
 Lab data vaginal discharge
- 3 Data collection - burning on urination
 Lab data burning urination
 treatment trichomonal vaginitis
 symptoms of trichomonal vaginitis

symptoms renal impaired hypertension - drug
 most causative organism for urinary tract
 infection
 age of occurrence of torsion of testicle
 bacteria associated with glomerulonephritis
 data collection for pressure on urination
 Infection
 Pain
 Dysfunction
 Renal stones
 Uretural colic

10

Neurology - Acute & Maintenance

Symptoms of meningeal irritation
 Data collection for sudden numbness right
 hand
 Cause for sudden numbness right hand
 cause of pounding headache - 6 hours
 cause of meningitis age 1 - 4
 symptoms of migraine
 gram stain meningitis shows
 TIA
 Aphasia - data
 Stroke
 Seizure disorders
 Paralysis dysfunction
 Vertigo

11&23

Endocrine - Acute & Chronic

treatment hypertensive patient with abnormal
 glucose tolerance test
 3 treatment diabetes
 8 Data collection diabetes
 Symptoms diabetes
 Information about insulin
 2 Data collection hyperthyroid
 2 Treatment hyperthyroid
 treatment obesity
 2 Complications diabetes
 2 Data collection diabetic acidosis
 Diagnosis diabetic acidosis
 Indeterminant sex in newborn
 Hypothyroidism
 Menopausal
 SLE
 Adrenal/Addisons Disease
 Identification adolescent male sexual
 development
 Followup for infant thymic irradiation

12

Dermatology Acute & Maintenance

treatment sudden rash on hands
 data collection rash arm & neck

2 year old with temporal lesion - data collection
 diagnosis herpes simplex
 diagnosis of pain following puncture wound
 areas affected in contact dermatitis
 dermatological sign of drug reactions
 treatment human bite
 diagnosis aphthous ulcers
 Erythema/allergy
 Impetigo
 Pityriasis rosea
 Infections
 Animal bite data & management
 Melanoma & other CA
 Psoriasis/eczema
 Rashes
 Pruritis

13

Eyes & ENT - Acute & Maintenance

edema following bacitracin ophthalmic ointment
 treatment 10 month old retracted tympanic membranes, fever . .
 diagnosis glaucoma
 cause of sore throat
 treatment corneal laceration
 diagnosis ear pain
 treatment ear pain with fever
 treatment otitis media
 cause otitis media
 cause hearing loss in children
 chronic treatment staph blepharitis
 cause of a chalazion
 Infectious allergy
 Foreign body
 Conjunctivitis management

14

Hematology - Acute & Maintenance

treatment acute blood transfusion reaction
 diagnosis generalized bleeding 7 year old
 cause nose bleed in alcoholic
 cause jaundice at birth
 treatment for heparin reaction
 dizziness spells in 14 year old
 symptoms leukemia in children
 iron deficiency anemia
 ecchymosis
 petechiae in children management & data
 Hemorrhagic purpura

15

Reproductive - Acute & Maintenance

- 2 diagnosis vaginal bleeding
 data collection
- 2 data collection acute low pelvic pain
 diagnosis bleeding, decreased blood pressure
 with IUD

conditions associated with vaginal discharge
 treatment vaginitis during pregnancy
 prescription birth control pills with family
 history thrombophlebitis
 treatment vaginal bleeding after menopause
 description of chancroid
 symptoms of eclampsia
 diagnosis uterine bleeding
 Amenorrhea
 Infectious Pain
 Vasectomy
 Cancer male & female
 Management positive serology
 Menstrual irregularities
 Pregnancy
 Testicular Torsion

16

Psychosocial - Acute & Maintenance

response to anxious mother
 data collection drug overdose
 2 follow-up care drug overdose
 treatment family disagreement in waiting room
 diagnosis psychiatric disorder
 data collection in irritability & difficulty
 sleeping
 follow-up for drug error
 response when 21 year old attempting to
 obtain drugs
 suspect adolescent may be pregnant & is
 afraid
 confidentiality of 19 year old who is
 pregnant
 mother questions quality of care
 data collection on resistive diabetic
 facts about child abuse
 Acute grief reaction
 Seductive patient management
 Suicide
 ETOH Intoxication
 Depression/anxiety reaction
 Obesity

17

Cardiovascular - Chronic

diagnosis loss of libido with hypertension
 data collection newly discovered murmur
 diagnosis murmur
 causes murmur
 follow-up murmur
 precipitating factors in angina
 Edema
 Hypertension
 Arterial/Venous Insufficiency Peripheral
 Cerebral Insufficiency
 Management Rheumatic fever

diagnosis inverted T wave
 cause of ankle edema
 data collection ankle edema
 symptoms CHF

18

Musculoskeletal - Chronic

data collection low back pain
 arthritis patient qualified for handicapped
 parking permit
 diagnosis hip injury
 management hip injury
 data collection bone cysts
 diagnosis serum calcium
 Arthritis
 Stroke
 Osteoarthritis - knee

19

Respiratory - Chronic

2 drugs used for TB
 PO₂ in asthma
 diagnosis non-productive cough
 COPD/emphysema
 Chronic bronchitis
 anti asthmatic prophylactic management

20

Gastrointestinal - Chronic

Indications for surgery in diverticular
 disease
 conditions thought to activate peptic ulcer
 disease
 Side effects of antacids
 treatment of progressive diarrhea
 data collection progressive diarrhea
 data collection hepatitis
 treatment chronic hepatitis
 treatment hemorrhoids
 Complications duodenal ulcers
 Occurrence of duodenal ulcers
 Diagnosis anemia and weight loss
 Neoplastic
 Ostomy Care
 Ulcer, gall bladder
 Obesity
 Constipation/Encopresis

21

Genitourinary - Chronic

2 Diagnosis renal failure
 Data collection recurrent vaginitis
 Treatment recurrent vaginitis
 Symptoms hydrocele
 Follow-up care cystitis
 Treatment condylomata acuminata
 Inguinal hernia
 Uterovesical prolapse

Chronic UTI
 Uremia
 Mechanical (Cystocele & Uterine prolapse)

- 22 Neurology - Chronic
 Diagnosis gradual personality change 78 year old
 2 Data collection "black out spells"
 Diagnosis "black out spells"
 Data collection multiple sclerosis-Weakness
 Diagnosis of dementia
 Hemiplegia
 Paralysis/Aphasia
 Headache
 Stroke
 Epilepsy - data & management
- 24 Dermatology - Chronic
 Symptoms Alopecia areata
 treatment Xerosis
 Pruritis
 Psoriasis
 Allergy
 Urticaria
- 25 Eyes & ENT - Chronic
 Diagnosis hearing loss
 treatment hearing loss
 diagnosis from audiogram
 diagnosis recurrent epistaxis
 diagnosis chronic nasal congestion
 symptoms - cancer larynx
 cause of "cauliflower ear"
 Serous O.M.
 Headache
 Sinusitis/Otitis
 Allergy
 Glaucoma
 Cataract
 Poor vision - data, management
 Vertigo
- 26 Hematology - Chronic
 symptoms iron deficiency anemia
 need for anemia screening
 data collection sickle cell anemia
 symptoms pernicious anemia and folic acid deficiency
 diagnosis pica
 Chronic anemia
 Leukemia
- 27 Reproductive - Chronic
 Counseling for abortion 16 year old

- treatment severe cramping
- 2 Data collection - difficulty becoming pregnant
- Identification benign breast lesion
- Menstrual disorders
- Pain
- Endometriosis
- Birth control
- Atrophic vaginitis

28

Psychosocial - Chronic

- Data collection - obesity
- Counseling - obesity
- 2 Data collection chronic abdominal pain
- Diagnosis chronic abdominal pain
- treatment chronic abdominal pain
- Counseling regarding side effects of chemotherapy
- Counseling regarding alcohol intake on job
- Counseling regarding breast mass
- 2 Data collection - depression
- Chronic organic brain syndrome
- Neurosis
- Family planning
- Adjustments social - economic - medical circumstances
- Substance abuse
- Anxiety
- Psychosis

Development of New Items

Data analysis has played an important part in improving the self-assessment exam. The items in the test, as it existed in the previous project, were evaluated in the following way. Item analyses were run on data gathered from two different samples of PAs who had taken the test as part of the previous project. One sample had been solicited by mail with 108 PAs completing and returning the test. The second sample consisted of 204 PAs who volunteered to take the SAE at the April 1980 Annual Convention of the AAPA. Items were flagged which had a negative discrimination index, an extremely high or low difficulty level, and an unused or scarcely used option. Additionally, scale homogeneity analyses were run, which provided the correlation of each of the items in a scale (Body System or Role) with that total scale score. Items with a correlation coefficient of less than 0.30 were marked. Based on these analyses, items were classified into three categories; those to be thrown out, kept, or revised.

As a crucial aspect of a criterion-referenced test is the representativeness of the items in a scale, this review was the first step of the refinement process, and was accomplished in conjunction with the project Advisory Committee.

The Advisory Committee was asked to review the list of critical incidents from which items in the test were taken. They were asked if the critical incidents were representative of PA practice. They were to refer back to the 11

major role areas, and note for each role area those minor responsibilities that did not seem to be covered in sufficient detail. For the Body System scales, they were asked for their judgement as to what important content areas were not covered. Their recommendations for content areas that needed to be represented by items within the existing scales were developed into a list. The project staff then contacted some available medical professionals, and asked them to write items covering these content areas using the critical incident technique. These new and revised items were assembled into a test, and mailed to 350 senior PA students in December, 1980. Item analyses were run, and items statistics were examined to isolate the better items. These items went into the final Form A4 of the test.

Scoring System Changes

At the scoring system level, three refinements recommended by the Advisory Committee were incorporated into the Self-Assessment Exam. These were error pattern scales, a weighted scoring system, and competency standards on each System and Role scale for two levels--entry and generalist. These additions all involved judgements by an expert Committee.

Weighted Scoring: The previous test had employed a dichotomous scoring system, such that the option chosen was either right or wrong. To improve the test in terms of providing

as much and as specific diagnostic feedback as possible, a weighted scoring system was implemented. By mail, the Committee was sent the items and options for the revised test and was asked to rate each option according to the following scale (4=effective, 3=acceptable, 2=unacceptable, 1=harmful). Their responses were compiled by the staff. The end result was an assignment of a weight (1,2,3, or 4) to each of the options for each item. Thus, with 320 items on the test, the total possible score under this weighted scoring system became 4×320 or 1280, rather than 320 under the dichotomous scoring system. The 17 Role Scales and the 29 System Scales are also scored using this weighted system. The reliabilities of the system scales using this weighted scoring system was compared to those using the dichotomous scoring system. In all but a few cases, the reliability of the weighted scoring system was equal to or greater than the reliability of the dichotomous scoring system (Table 5). The reliability estimate used was that of Cronbach's alpha.

Competency Standards: In order to give the test-taker an idea of how he or she stands relative to the domain of content, a standard or cut-off score must be established for competency. Previously, a standard was set for each system and role scale. To improve the test in terms of providing as accurate diagnostic feedback as possible, standards for each scale were set for 2 different levels of PAs-the entry level and the advanced/experienced level. Entry-level de-

TABLE 5

RELIABILITY OF SYSTEM SCALES: WEIGHTED VERSUS DICHOTOMOUS
SCORING

System Scale	Weighted Scoring System	Dichotomous Scoring System
1 Cardiovascular-Emergency	.394	.207
2 Respiratory-Emergency	.314	.279
3 Gastrointestinal-Emergency	.296	.293
4 Neurology-Emergency	.213	.101
5 Cardiovascular-Acute	.434	.407
6 Musculo-Skeletal-Acute	.276	.267
7 Respiratory-Acute	.419	.427
8 Gastrointestinal-Acute	.199	.151
9 Genitourinary-Acute	.091	.081
10 Neurology-Acute	.296	.275
11 Endocrine-Acute	.098	.066
12 Dermatology-Acute	.239	.182
13 Eyes, ENT	.339	.327
14 Hematology	.290	.250
15 Reproduction	.256	.234
16 Psychosocial	.139	.093
17 Cardiovascular-Chronic	.301	.276
18 Musculo-Skeletal-Chronic	.393	.377
19 Respiratory-Chronic	.096	.046
20 Gastrointestinal-Chronic	.302	.267
21 Genitourinary-Chronic	.146	.181
22 Neurology-Chronic	.189	.113
23 Endocrine-Chronic	.303	.296
24 Dermatology-Chronic	.336	.190
25 Eyes, ENT	.207	.149
26 Hematology	.335	.223
27 Reproductive	.273	.186
28 Psychosocial	.244	.224
29 Other	.304	.185

lines those PAs with less than five years' experience, and advanced-level defines those PAs with five years of experience or more. A modified version of the Nedelsky method (1954) was used to arrive at these standards. This method involves a review of the options in each item to identify those options that would be rejected by a minimally competent PA. Nedelsky believed that a set of judges could make such decisions reasonably consistently, and thus, come up with a dependable minimum competency score.

By mail, the Advisory Committee was provided with a list of the set of new, old, and revised items. They were asked to identify for the entry and advanced level of PA those incorrect options for each item that a minimally competent PA at each level may be expected to eliminate. The Committee's responses were compiled and a modified Delphi technique was utilized to arrive at consensus. An expected standard or cut-off score for each item for each of the two levels of PA experience--entry and experienced--was then calculated in the following way: If the Committee consensus judged an entry level PA could eliminate only one option out of five, then the expected score would be one out of four (the number of incorrect options), or 25%. Then this percentage was multiplied by the average weight of the options identified as those that a minimally competent PA could not eliminate as incorrect. The weight of each option was that assigned to it as part of the weighted scoring system described previously, e.g., 4=effective, 3=accept-

able, 2=unacceptable, 1=harmful. For each scale, a standard was calculated by summing the expected scores for the items on that scale for each level of PA: entry and advanced. Thus, whereas the original minimum competency scores from the previous project had been calculated by asking the judges to identify options for each item that would be rejected by a minimally competent entry level PA, the refinement involved asking the judges to go through the process twice; once for minimally competent entry level PAs and once for minimally competent experienced PAs.

Error Pattern Scales: The third type of information or feedback that was incorporated into the revised test during this project was error pattern scale scores. In the first project, a list of types of errors commonly made when answering multiple-choice items was generated and supplied to the item writers who were asked to write the incorrect options for the items keeping these error categories in mind. In this project, this categorization of types of errors was expanded incorporating suggestions provided by the new item writers and the Advisory Committee, who made judgements as to the assignment of options to error categories.

After choosing the items that would go into Form A4, each of the options for this revised set of items needed to be classified as belonging to a certain type of error, so that error pattern scales would be defined. To accomplish

this, the Committee was mailed a copy of the items with the multiple-choice options listed down the side and the five error categories, taken from the work of the previous project, listed along the top of the page. They were asked to assess the options for each item in terms of the type of error pattern they represented, adding other error categories besides the five given, as they saw fit. These data were compiled by the staff and the options were assigned to one of thirteen error scales based on the Committee's inputs. This assignment was the basis for the error pattern scales which were incorporated into the score reporting system. The 13 Error Pattern Scales used in the exam are shown in Table 6.

3. Reliability and Validity

"The main concern of test theory is the examination of the measurement process in order to be able to indicate how well a given procedure serves some purpose" (Th. van der Kamp, 1976). In order for the SAE to be used confidently evidence must be presented that the scores reported are reliable and valid. The validity of a test refers to whether or not test performance can be used for the purpose for which it was designed (i.e., to predict academic success, to determine competency, to describe a person's standing on a personality dimension). Whether or not a test is valid for the use demanded of it is an empirical question. Reliabil-

TABLE 6
THE 13 ERROR PATTERN SCALES

1. Insufficient Knowledge
2. Insufficient Data Gathering
3. Inaccurate Data Analysis
4. Premature Diagnosis
5. Inaccurate Diagnosis
6. Incorrect Treatment
7. Ineffective Treatment
8. Delayed Treatment
9. Inappropriate Drug Use
10. Inappropriate Referral
11. Ineffective Interpersonal
12. Inappropriate Interpersonal
13. Excessive Data Gathering

ity refers to evidence about the consistency of scores over time or forms. Typically, reliability coefficients describe the extent to which the variance of the scores is due to systematic variance rather than error variance.

Because of the extensive information reported in this exam, we have only barely touched the surface of examining the reliability and validity of each of the different types of scores that are reported. Scores are presented for 17 role subscales, 29 body system subscales, and 13 error pattern subscales, referenced to an absolute standard of competence. Scores are also presented on the 29 system scales in a norm-referenced format in the form of stanines. Evidence needs to be presented on the reliability and validity of these scale scores and of the cut-off scores established for each subscale. Because the profile is based on a norm-referenced interpretation, the characteristics of the norm group should also be described. Because the objective of this project was the refinement of the exam, which meant more test development than test evaluation, many of the reliability and validity studies have yet to be designed. We will present what analyses we have done, and suggest what studies may still need to be done.

Reliability

We will begin by examining evidence regarding the reliability of the subscale scores. Because scale scores are derived from a collection of items the variance across the

items should be systematic. That is, examinees should order themselves similarly across the items in a scale measuring a construct. A measure of this internal consistency of items is typically computed by using the formula for Cronbach's Alpha. In essence, each item is considered a form of the test, and Alpha is an estimate of the degree to which these forms or items go together. Alpha for each of the 17 role scales and 29 system scales are presented in Tables 7 and 8.

These reliabilities are fairly low indicating that the variance due to difference among people is not accounting for a great deal of the observed variance. Part of the problem may be that many of the scales have few items, which lowers reliability. However, a major part of the problem lies in the criterion-referenced nature of the exam. In other words, the items were selected to reflect a certain criterion/standard and are not intended to create variance or differences in scores obtained by subjects. The power of the exam was designed to lie in differentiating masters versus non-masters as related to entry-level competence in the PA profession.

It has been argued that internal consistency estimates of reliability, as reported in the two tables above, are inappropriate for criterion-referenced tests (Popham and Husek, 1969) because the tests are not constructed to maximize variance upon which internal consistency estimates depend. If it is expected that all examinees will pass an item, then there will be no variance on the item, and across

TABLE 7

Role Scales: Reliability Indices

Role Scale	# items	Cronbach's Alpha N=1283	Coefficient of Acceptance N=891
1 Professional Role	5	.165	.91
2 Interpersonal Behavior	14	.226	.92
3 Gather Data-Resp,CV, EENT	18	.431	.73
4 Gather Data-GI, GU	15	.187	.95
5 Gather Data-Psych., Neuro	29	.416	.59
6 Analyze Data-Endocrine	18	.342	.61
7 Analyze Data-Resp,CV, EENT	30	.555	.67
8 Analyze Data-Hematology	13	.421	.75
9 Analyze Data-GI-GU	23	.331	.69
10 Analyze Data-Psych. Neuro.	17	.381	.52
11 Formulate Plan-Mus. Skel.	23	.414	.62
12 Handle Dermatology	18	.437	.80
13 Manage Patients-Resp.CV, EENT	30	.423	.69
14 Manage Patients-GI, GU	16	.221	.78
15 Manage Patients-Reproductive	5	.274	.77
16 Manage Patients-Psych. Neuro.	10	.089	.79
17 Handle Pharmacology	20	.346	.57

TABLE 8
SYSTEM SCALES: RELIABILITY INDICES

System Scale	# items	Cronbach's Alpha N=1283	Coefficient of Acceptance N=891
1 Cardiovascular- Emergency	14	.394	.90
2 Respiratory- Emergency	9	.314	.92
3 Gastrointestinal- Emergency	8	.296	.94
4 Neurology- Emergency	8	.213	.62
5 Cardiovascular- Acute	17	.434	.92
6 Musculo-Skeletal- Acute	14	.276	.66
7 Respiratory- Acute	16	.419	.56
8 Gastrointestinal- Acute	16	.199	.75
9 Genitourinary- Acute	13	.091	.53
10 Neurology- Acute	12	.296	.90
11 Endocrine- Acute	4	.098	.85
12 Dermatology- Acute	12	.239	.60
13 Eyes, ENT	12	.339	.77
14 Hematology	8	.290	.89
15 Reproduction	12	.256	.56
16 Psychosocial	13	.139	.71
17 Cardiovascular- Chronic	12	.301	.73
18 Musculo-Skeletal- Chronic	10	.393	.78
19 Respiratory- Chronic	10	.096	.50
20 Gastrointestinal- Chronic	12	.302	.77
21 Genitourinary- Chronic	8	.146	.84
22 Neurology- Chronic	9	.189	.72
23 Endocrine- Chronic	14	.303	.62
24 Dermatology- Chronic	6	.336	.95
25 Eyes, ENT	11	.207	.59
26 Hematology	7	.335	.67
27 Reproductive	5	.273	.60
28 Psychosocial	12	.244	.68
29 Other	16	.304	.71

TABLE 9
System Scales: Validity Indices

System Scale	Correlation with practice N=1283	Correlation with felt needs N=1283
*1 Cardiovascular- Emergency	.13	.20
*2 Respiratory- Emergency	.10	.10
3 Gastrointestinal- Emergency	.006	.07
4 Neurology- Emergency	.018	.02
*5 Cardiovascular- Acute	.08	.20
*6 Musculo-Skeletal- Acute	.19	.21
*7 Respiratory- Acute	.08	.20
8 Gastrointestinal- Acute	.09	.04
9 Genitourinary- Acute	-.06	-.006
10 Neurology- Acute	.02	.04
11 Endocrine- Acute	.02	.06
*12 Dermatology- Acute	.16	.21
*13 Eyes, ENT	.11	.11
14 Hematology	.003	.02
*15 Reproduction	.09	.11
16 Psychosocial	-.008	.04
*17 Cardiovascular- Chronic	.13	.21
18 Musculo-Skeletal- Chronic	.02	.01
19 Respiratory Chronic	.02	.04
20 Gastrointestinal- Chronic	.04	.06
21 Genitourinary- Chronic	.01	.06
22 Neurology- Chronic	.04	-.007
*23 Endocrine- Chronic	.11	.14
*24 Dermatology- Chronic	.09	.12
*25 Eyes, ENT	.10	.07
*26 Hematology	.11	.13
*27 Reproductive	.15	.17
*28 Psychosocial	.09	.11

items, an internal consistency estimate will be low. It is not that a criterion-referenced test cannot have high internal consistency, but if it does not, it may not necessarily mean that the items in the scale do not go together, or are not measuring the same thing.

The topic of appropriate measures of reliability for criterion-referenced tests has been discussed extensively in the measurement literature. Many have pointed out that the reliability of the criterion-referenced scores may not be as important as the reliability of the mastery classification decisions. That is, whether or not examinees' responses are consistent across items may not be as important in terms of effects on examinees, as whether or not examinees are or would be consistently classified as masters or non-masters.

For assessing the reliability of categories of mastery, several approaches have been suggested: 1) Carver (1970) 2) Swaminathan, Hambleton, and Algina (1974) 3) Huynh (1976) 4) Subkoviak (1976). Subkoviak (1980) compares the results of these different approaches. He concludes that the advantages and disadvantages of the Huynh and Subkoviak methods are generally similar. All three methods require only one test administration and result in estimates with relatively small standard errors for classroom size samples. The disadvantages of the three approaches are that they are tedious to compute and produce biased estimates for short tests. Subkoviak's method was used to calculate the reliability of

the SAE decision classification for the 17 Role Scales and the 29 System Scales on a partial sample (N=891) of the total sample of PAs who have taken the SAE. These estimates are included in Tables 7 and 8.

However, it may be premature to interpret these consistency classification estimates. According to Berk (1980), it may be pointless to even compute such a reliability index unless the cut-off score has been validated. "A high reliability index based on an invalid standard, for example, might indicate that a test can consistently classify students incorrectly across parallel forms" (p. 47). The cut-off scores for the 17 Role Scales and the 29 System Scales have not yet been validated empirically. Typically, this validation is accomplished by administering the test to two groups (i.e., an uninstructed and an instructed group), and observing whether or not the instructed group has a significantly higher percentage of examinees passing the cut-off. In the case of the PA-SAE, there are no easily available criterion groups to use in validating the cut-offs. It is possible that if the standards have been set too high and that examinees are being consistently classified as below competence, that a high reliability index would be achieved. This possibility needs to be ruled out by further studies.

Content Validity

There are three kinds of validity that are relevant when trying to provide evidence about the goodness of a test: construct, criterion or decision, and content. Content validity has been of primary concern in the majority of criterion-referenced tests, but Messick, (1975) and others have argued that although content validity is important for establishing content relevance, it does not provide evidence for the practical utilization and interpretation of scores. Content validity reflects judgements as to how closely items match their objectives or domain specifications and how well the items mirror the domain.

The work that went into the domain classification of the PA-SAE is relevant to content validity. (See Table 4 for list specifying critical incidents that must be represented in exam.) Any time that the domain is outlined and this outline is used to generate items, the beginnings of content validity are present. In order to further ensure content validity, it is necessary once the items are generated to have a Committee review their match or congruence to the scale to which they were assigned, and their representativeness in adequately covering the scale content area. This review took place at the first meeting of the Advisory Committee. The Committee reviewed the list of critical incidents as they existed from the previous project's exam development efforts, and suggested additions that they thought would improve the representativeness of the exam. New items

were then written to cover these critical incidents or content areas.

Construct Validity

Construct validity is the process of collecting evidence about expected empirical relations to support the inference that an observed response consistency has a particular meaning. In this case, we need to explore the extent to which the 17 role scales, the 29 body system scales, and the 13 error pattern scales produce scores that can be interpreted as reflecting a person's standing on the construct. Thus, a low score on a scale (cardiology, for example) needs to be shown to reflect a person's true level of competence in this area. A question to ask is, do people that score low on a scale exhibit some other behavior that validates this low scale score?

Two sets of data available from the self-assessment process that are relevant here are the practice profile questions and the expressed need for CME questions. Both of these sets of questions are keyed to each of the 29 system scale areas, but not for the role scale areas. So the evidence we have is only for system scales, not role scales or error pattern scales.

It was predicted that subjects who saw more patients in a certain system area (as indicated by the practice question) would have higher scores than those who saw less patients in this area because they would be exposed to the

critical incidents in that area more often. Thus, we expected positive correlations between each practice question and scale score for each of the 29 scales. It was also predicted that PAs who said they were most comfortable in a given area should have higher scores on the system scale area. The following results were obtained. The same pattern of relationships was found for both sets of correlations. System scales 1,2,5,6,7,12,13,15,17,23,24,25,26,27, and 28 had significant correlations between both scale scores and practice questions and scale scores and need questions. Thus, people who got higher scale scores on these scale areas tended to see more patients and feel more comfortable in these areas. These system scales also tend to be the ones with higher reliabilities (internal consistency) as can be seen from Tables 8 and 9. The starred System scales in Table 9 are those with all three types of evidence pointing towards construct validity.

The seeking of correlations among exam scores(E), practice profile(P), and expressed need(N) is not without theoretical problems. A research study exploring these relationships is underway as part of a Ph.D. thesis (Odor, 1982). A preliminary research paper (D'Costa and Odor, 1982) using log-linear analyses, concluded that extent of practice is correlated with exam performance in each System area.

Another study (D'Costa and Lower, 1982, in progress), compared supervisor ratings and exam scores with self-ex-

pressed needs identified on 3 occasions: prior to taking the exam, upon taking the exam, and after receiving feedback. It was found that the feedback does make a difference in terms of the validity of self-assessment. This study has not yet been completed nor reported in the literature. Although based on a small sample ($N=105$) and possibly biased (about 40% return), the data is expected to provide further evidence on the predictive and construct validity of the SAE.

Decision Validity

The scores on these scales are used in a particular way; to provide PAs with feedback about their competence. An individual's score is compared to a standard or absolute cut-off score for each scale which determines his or her competence or non-competence. In other words, the scores on each scale are being used to make decisions about the PA's classification and to encourage individual decisions for continuing competency maintenance. It is crucial that the validity of the cut-off scores used for classification purposes be verified (Hambleton, 1980). Ideally, we would like to set the standards and then collect evidence as to the validity of this standard by giving the test to 2 or more criterion groups who should differ in the proportion classified as masters or non-masters. This kind of evidence has not yet been collected, and it may be difficult to

collect as there are no clear criterion groups (such as instructed versus uninstructed PAs) to use.

However, other approaches are possible, although not yet attempted in this project period. Cross-validation, using canonical correlation techniques, is a distinct possibility. A Ph.D. thesis (Odor, 1982) is attempting to validate a CME model in which expressed need is the multivariate dependent variable, and exam scores, practice profile, and other biographical and professional characteristics are the independent variables. In addition to canonical analyses, a path analysis using multiple regression is being attempted in order to cast light on the validity of decisions made using self-assessment approaches.

Several other cross-validation techniques are also becoming popular for criterion-referenced tests in which part-samples are set aside to assess stability of decisions. Likewise item-analyses, using masters vs non-masters as criterion, rather than the usual total score based point biserial correlation, are coming into vogue. The computer programs for these techniques are not available and hence these analyses belong to the future.

4. Exam Reporting System

An eight-page Individual Report is developed and sent to each PA who takes the self-assessment. Each page is described in the Interpretive Booklet which accompanies the

Individual Report. It is expected that the PA will spend the time to study the Individual Report with the help of the Interpretive Booklet, and then, pursue the prescribed learning materials in order to remediate or enhance competence.

The Individual Report begins with a letter to the PA on Page One and provides more general information. With each new page additional details are given. The pages of the Report will now be described in some detail. See Figure 6.

Page One: Letter with Performance Summary and CME Recommendations

This page is a summary of the PA's performance relative to the standards set for each of the Role scales, System scales, and Error Pattern scales. Their performance on each of these scales is classified into Areas of Strength and Areas Needing Improvement. Strength is defined as scoring above the standard or cut-off score set for experienced or advanced-level PAs. Needing improvement is defined as scoring below the standard or cut-off set for Entry level PAs. Cut-off scores were set by expert judgement.

Scales reported under Areas of Strength and Areas Needing Improvement are rank ordered in terms of strength or need.

For each strength and deficiency a learning prescription is recommended. Currently, the system does not provide learning prescriptions for error patterns because appropriate learning modules not yet available.

Page Two: Practice, Need Expressed, and Exam Performance Profile

Perhaps the most useful and clearly unique feature of the report is this page which provides a combined picture for the PA of his Practice profile (the number of patients seen in the PA's practice), the extent of discomfort expressed by the PA in terms of wanting CME in the specific area, and the exam score earned for each of the 29 System Scales. In order to make comparisons across the three types of scores, a normative reference system was used. A stanine rating is computed for each of the three types of scores. Stanines are based on a simple linear transformation of scores so the new mean and standard deviation are always 5 and .5 respectively. All the indicators can be compared with one another because they have been transformed into this stanine system. This capability to compare indicators is the crux of the P,N,E report. The examinee's stanine level can be from 1 to 9 and is coded such that 2 dots/dashes/asterisks are equal to one stanine. Thus, the report presents 3 lines of dots/dashes/asterisks representing the person's level on practice, needs, and exam scores. The respondee is advised to consider this profile to discover more about their needs.

Page Three: Actual and Correct Responses to Exam Items

As shown in Figure 6, this page provides a straightforward list of responses made by the PA to each item in the exam. It is assumed that the exam will be available

to the PA for review purposes. If the response is wrong, the correct response is indicated within parentheses. Responses judged "harmful" to a patient, i.e. scored with weight of 1, are shown with asterisks. The occurrence of several asterisks on this page serves as a red flag to the PA.

Page Four: Role Scales: Raw Scores Relative to Competency Standards

The seventeen Role Scales are listed and the PA's performance on each scale is shown graphically on a straight line which has been standardized to take care of scales with unequal number of items. Overlaid on the line is a competency band (asterisks) with the Entry-level standard at one end and the Advanced level standard at the other. The PA's own score is shown by the symbol S. If S falls below the competency band, the PA is judged deficient in this scale and a negative (-) sign is shown in parentheses by the scale name. The scale is also listed on Page One of the Individual Report as a deficiency. If S falls above the competency band, the PA is judged to have strength (shown by +) in this scale area. If S falls on the competency band, no sign appears and the PA is judged to have acceptable competence in this area.

Page Five: Role Scales: Quality of Responses

The seventeen Role Scales are again listed but this time the PA's responses are broken down to show how many were judged harmful, unacceptable, acceptable, and effec-

tive. This allows the individual to find out the seriousness of deficiency/strength in each scale. Since the "keys" to the Scales are listed in the Interpretive Booklet, the PA can go over each item in the scale and determine where his/her knowledge strengths/deficiencies are. The detailed information provided here helps to mitigate in situations where the response was not the "correct" or "effective" one, but somewhat less. This information also serves an educational purpose.

Page Six: System Scales: Raw Scores Relative to Competency Standards

This page provides a similar print-out as Page Four, except that the feedback relates to the System scales. Receiving competency standings relative to two approaches to scoring, the Role approach and the Body System approach, helps to specify the PA's strengths and/or weaknesses better. The two approaches are somewhat different in their approach to competency, one focuses more on daily role responsibilities and the other on clinical knowledge/skills needed.

Page Seven: System Scales: Quality of Responses

This page is similar to Page Five. The feedback now relates to the System scales rather than Role scales.

Page Eight: Error Pattern Scales: Raw Scores Relative to Competency Standards

This page is similar to Pages Four and Six, except that this time the feedback relates to Error Patterns. The

competency band is to be interpreted differently since the Entry-level Standard is at the opposite end. Entry-level persons would be expected to make more errors than Advanced-level persons.

Error Patterns were established on the basis of responses identified to be less than "effective". Potentially, an error can be committed when an option other than "effective" is selected. However because of the large number of possible error keys, selectivity was used so that only the most critical errors were keyed.

Interpretive Booklet

A 28 page Interpretive Booklet is provided to each PA who takes the self-assessment exam with the Individual Report. This Interpretive Booklet provides an introductory explanation of the exam and its technical development. It also provided assistance on how each page of the Report can be utilized for one's professional development. Provided as an Appendix to the Interpretive Booklet is a List of Learning Resources with four letter codes. These codes enable the PA to identify the specific learning prescription indicated in the Individual Report (Page One).

AMERICAN ACADEMY OF PHYSICIAN ASSISTANTS
SELF-ASSESSMENT EXAMINATION
PA ID#

MAY 1981

DEAR PHYSICIAN ASSISTANT:

THIS INDIVIDUAL REPORT FORM SPECIFIES AND ANALYZES YOUR PERFORMANCE ON THE SELF-ASSESSMENT EXAMINATION. YOUR SCORES ARE PRESENTED IN TERMS OF 17 ROLE SCALES, 29 SYSTEM SCALES, AND 13 ERROR PATTERN SCALES. PLEASE REFER TO THE INTERPRETIVE BOOKLET FOR A DETAILED EXPLANATION OF THE FORMAT AND CONTENT OF THE REPORT FORM. YOU ARE ALSO ENCOURAGED TO PLAN YOUR CME RELATIVE TO YOUR PRACTICE DEMANDS, YOUR PERSONAL NEEDS, AND YOUR EXAM PERFORMANCE. SOME SUGGESTIONS ARE LISTED BELOW:

RECOMMENDATIONS FOR CONTINUING MEDICAL EDUCATION
AND PROFESSIONAL DEVELOPMENT

AREA(S) OF STRENGTH

RESOURCES RECOMMENDED

ROLE SCALES

SYSTEM SCALES

5 ACU CARDIOVASCULAR
2 EMR RESPIRATORY

FA01	FB	FC01	FD01
EA01	EB	EC	ED01

ERROR PATTERN SCALES

12 INAPPROP INTERPERSONAL
3 INACCURATE DATA ANALYSIS
10 INAPPROPRIATE REFERRAL

AREA(S) NEEDING IMPROVEMENT

ROLE SCALES

17 HANDLE PHARMACOLOGY PROB
16 MANAGE PTS-P SYCH,NEURO
5 GATHER DATA-PSYCH,NEURO

SYSTEM SCALES

19 CHR RESPIRATORY
29 OTHER-(ECON,DIET)
23 CHR ENDOCRINE
28 CHR PSYCHOSOCIAL
9 ACU GENITOURINARY
25 CHR EYES & ENT
4 EMR NEUROLOGY
27 CHR REPRODUCTIVE
7 ACU RESPIRATORY
12 ACU DERMATOLOGY

EA01	EB	EC	ED01
CA01	CB		CD01
LA01	LB	LC01	LD01
IA01	IB		ID01
DA	DB		DD01
KA01	KB		KD01
JA01	JB		JD01
EA01	EB	EC	ED01
BA01	BB		BD01

ERROR PATTERN SCALES

CONTINUED TO PAGE 2

PROFILE OF PRACTICE, NEEDS, AND EXAM SCORES
PAGE 2 OF REPORT FOR AAPA ID NUMBER

COMPARED WITH YOUR PRACTICE (P) PROFILE (TYPES OF PATIENTS SEEN)
AND WITH YOUR CME NEEDS (N) PROFILE (AREAS SOUGHT TO LEARN)
YOUR EXAM (E) SCORES WOULD SHOW UP AS FOLLOWS:

SYSTEM SCALES

EMERGENCY

1 CARDIOVASCULAR P.....
N.....
E*****

2 RESPIRATORY P.....
N.....
E*****

3 GASTROINTESTINAL P.....
N.....
E*****

4 NEUROLOGY P.....
N.....
E*****

ACUTE

5 CARDIOVASCULAR P.....
N.....
E*****

6 MUSCULOSKELETAL P.....
N.....
E*****

7 RESPIRATORY P.....
N.....
E*****

8 GASTROINTESTINAL P.....
N.....
E*****

9 GENITOURINARY P.....
N.....
E*****

10 NEUROLOGY P.....
N.....
E*****

11 ENDOCRINE P.....
N.....
E*****

12 DERMATOLOGY P.....
N.....
E*****

13 EYES & ENT P.....
N.....
E*****

14 HEMATOLOGY P.....
N.....
E*****

15 REPRODUCTIVE P.....
N.....
E*****

16 PSYCHOSOCIAL P.....
N.....
E*****

CHRONIC

17 CARDIOVASCULAR P.....
N.....
E*****

18 MUSCULOSKELETAL P.....
N.....
E*****

19 RESPIRATORY P.....
N.....
E*****

20 GASTROINTESTINAL P.....
N.....
E*****

21 GENITOURINARY P.....
N.....
E*****

22 NEUROLOGY P.....
N.....
E*****

23 ENDOCRINE P.....
N.....
E*****

24 DERMATOLOGY P.....
N.....
E*****

25 EYES & ENT P.....
N.....
E*****

26 HEMATOLOGY P.....
N.....
E*****

27 REPRODUCTIVE P.....
N.....
E*****

28 PSYCHOSOCIAL P.....
N.....
E*****

29 OTHER-(ECON,DIET) P.....
N.....
E*****

30 P..
N..
E**

INDIVIDUAL AND CORRECT RESPONSES
PAGE 3 OF REPORT FOR AAPA ID NUMBER

YOUR TOTAL WEIGHTED SCORE IS 1039 OUT OF 1280;
TOTAL UNWEIGHTED SCORE IS 196 OUT OF 320;
YOU OMITTED 1 ITEMS.
THE ENTRY LEVEL MINIMUM COMPETENCY SCORE IS 1030.

YOUR RESPONSES TO ITEMS ARE PRINTED BELOW. THE CORRECT
RESPONSE IS SHOWN WITHIN PARENTHESES ONLY WHEN DIFFERENT.
AN ASTERISK PRECEDING YOUR RESPONSE INDICATES THAT YOU
SELECTED AN OPTION JUDGED 'HARMFUL' TO THE PATIENT.

SECTION 1

1 E	11 A	21 B	31 E(A)	41*(D)	51 E	61 D	71 A
2 E	12 B(C)	22 D	32 D	42 A(B)	52 C(E)	62 C	72 C
3*(A)	13 D(C)	23 D	33 E(D)	43 C	53 D(A)	63 B(C)	73 B(A)
4 D	14 C(E)	24 A	34 A(C)	44 A	54 C	64 C(D)	74 C
5 A(C)	15 A	25 C(E)	35 C(D)	45 B(A)	55 B(E)	65 B	75 A
6 B	16 E(C)	26 A(D)	36 A	46 D(E)	56 B(E)	66 B	76 E
7 D	17 B(D)	27 E	37 A(D)	47 E(A)	57 B(A)	67 B	77 D(A)
8 B(D)	18 B(D)	28 C	38 E	48 C	58 C	68 B(A)	78 A(C)
9 C	19 A(B)	29 C	39 C(E)	49 D	59 C(D)	69 A	79 E
10 D	20 D	30 D(B)	40 B(A)	50 D	60 D	70 D	80 C(B)
81 A(B)	91 C(D)	101 C(E)	111 D	121 C(B)	131 A	141 C(A)	151 A
82 D(C)	92 A	102 C	112 C	122 D	132 C	142 A	152 D
83 C(A)	93 A(C)	103 C	113 B	123 E	133 B	143 D	153 C
84 A(D)	94 A	104 D	114 B	124 C(B)	134 B(A)	144 A	154 B(D)
85*(A)	95 A	105 B	115 B	125 C	135 C	145 A	155 B
86 E	96 D	106 B	116 A	126 C(A)	136 C(D)	146 B	156 C(D)
87 A(B)	97 A(D)	107 B(D)	117 A	127 A	137 B	147*(D)	157 A(B)
88 B(C)	98 B	108 D(B)	118*(D)	128 C	138 A	148 A	158 D(E)
89 D	99 C(B)	109 D	119 E(C)	129 E(D)	139 C	149 B(D)	159 C
90 A(E)	100 C(A)	110 E	120 A(D)	130 C	140 B(E)	150 B(E)	160 D

SECTION 2

1 E	11 E	21 C(E)	31 C	41 C(B)	51 B(A)	61 C	71 B
2 B	12 C	22 B(D)	32 A(B)	42 D(C)	52 C	62 C	72 B
3 D	13 C(B)	23 C	33 D	43 B(A)	53 E	63 A	73 D
4 D(A)	14 C	24 B	34 B(A)	44 B(A)	54 D	64 C	74 C
5 C	15 B	25 B	35 E(D)	45 C	55 C(B)	65 C(B)	75 B(E)
6 B	16 D	26 D(B)	36 E	46 C	56 D	66 A(B)	76 D(A)
7 A(C)	17 D(C)	27 D	37 C	47 C	57 A(B)	67 D	77 C
8 C	18 C	28 A	38 A(B)	48 C(E)	58 B	68 A(D)	78 B(D)
9 C	19 C	29 D	39 E	49 B	59 E	69 A	79 D(C)
10 C	20 D	30 D	40 A	50 A(D)	60 C	70 A	80 D(A)
81 D(B)	91 B	101 A(C)	111 E	121 B	131 C(D)	141 E	151 C
82 C	92 E	102 B	112 C	122 D	132 B(C)	142 B	152 C(D)
83 C	93 B(A)	103 B	113 B	123 E	133 B	143 D(E)	153 C
84 A	94 B	104 C	114 C	124 D	134 B(D)	144*(A)	154 C
85 A	95 E	105 E	115 C(A)	125 C	135 E	145 B	155 C
86 B	96 A	106 C	116 A	126 C(A)	136 C	146 D(E)	156 D
87 B	97 B(C)	107 C	117 B	127 D(A)	137*(D)	147 A(C)	157 B(E)
88 B	98 B	108 B	118 E	128 C(B)	138 B	148 B	158 C(E)
89 C	99 A	109 C	119 C(A)	129 C	139 D	149 B	159 C
90 A(E)	100 D	110 D	120 B	130 C(B)	140 E	150 B(A)	160 C

CONTINUED ON PAGE 4

WEIGHTED SCORES RELATIVE TO ROLE SCALE COMPETENCY BANDS
PAGE 4 OF REPORT FOR AAPA ID NUMBER

IN TERMS OF ROLE SCALES, YOUR WEIGHTED SCORES (S) IN RELATION TO A COMPETENCY BAND (*****) APPEAR BELOW.
A MINUS OR PLUS SYMBOL BEFORE A SCALE NUMBER INDICATES THAT YOUR SCORE IS BELOW OR ABOVE THE COMPETENCY BAND.

ROLE SCALES	YOUR SCORE	MAX SCORE	
() 1 PROFESSIONAL ROLE	16	20	*S*.....
() 2 INTERPERSONAL BEHAVIOR	45	56	*S****..
() 3 GATHER DATA-RESP, CV, EENT	61	72	**S**....
() 4 GATHER DATA-GI, GU	50	60	***S***..
(-) 5 GATHER DATA-PSYCH, NEURO	95	116	S****..
() 6 ANALYZE DATA-ENDOCRINE	59	72	S***....
() 7 ANALYZE DATA-RESP, CV, EENT	106	120	***S***..
() 8 ANALYZE DATA-HEMATOLOGY	40	52	*S***....
() 9 ANALYZE DATA-GI, GU	73	92	S*****..
() 10 ANALYZE DATA-PSYCH, NEURO	60	68	***S*....
() 11 FORMULATE PLAN-MUSC/SKEL	75	92	S****..
() 12 HANDLE DERMATOLOGY PROB	55	72	S****....
() 13 MANAGE PTS-RESP, CV, EENT	98	120	*S****..
() 14 MANAGE PTS-GI, GU	56	64	***S*....
(+) 15 MANAGE PTS-REPRODUCTIVE	20	20	***..S
(-) 16 MANAGE PTS-PSYCH, NEURO	31	40	S****....
(-) 17 HANDLE PHARMACOLOGY PROB	54	80	S.*****..

CONTINUED ON PAGE 5

QUALITY OF ROLE SCALE RESPONSES
PAGE 5 OF REPORT FOR AAPA ID NUMBER

IN TERMS OF ROLE SCALES, THE QUALITY OF YOUR RESPONSES APPEARS AS FOLLOWS:

H = HARMFUL U = UNACCEPTABLE
A = ACCEPTABLE E = EFFECTIVE

ROLE SCALES	MAXIMUM NUMBER OF RESPONSES	QUALITY OF RESPONSES			
		H	U	A	E
1 PROFESSIONAL ROLE	5	0	2	0	3
2 INTERPERSONAL BEHAVIOR	14	0	5	1	8
3 GATHER DATA-RESP, CV, EENT	18	1	4	0	13
4 GATHER DATA-GI, GU	15	0	5	0	10
5 GATHER DATA-PSYCH, NEURO	29	0	10	1	18
6 ANALYZE DATA-ENDOCRINE	18	0	6	1	11
7 ANALYZE DATA-RESP, CV, EENT	30	0	6	2	22
8 ANALYZE DATA-HEMATOLOGY	13	0	6	0	7
9 ANALYZE DATA-GI, GU	23	0	9	1	13
10 ANALYZE DATA-PSYCH, NEURO	17	0	3	2	12
11 FORMULATE PLAN-MUSC/SKEL	23	1	7	0	15
12 HANDLE DERMATOLOGY PROB	18	0	6	1	10
13 MANAGE PTS-RESP, CV, EENT	30	1	9	1	19
14 MANAGE PTS-GI, GU	16	0	3	2	11
15 MANAGE PTS-REPRODUCTIVE	5	0	0	0	5
16 MANAGE PTS-PSYCH, NEURO	10	1	2	2	5
17 HANDLE PHARMACOLOGY PROB	20	2	10	0	8

CONTINUED ON PAGE 6

WEIGHTED SCORES RELATIVE TO SYSTEM SCALE COMPETENCY BANDS
PAGE 6 OF REPORT FOR AAPA ID NUMBER

IN TERMS OF SYSTEM SCALES, YOUR WEIGHTED SCORES (S) IN RELATION TO
A COMPETENCY BAND (*****) APPEAR AS FOLLOWS:

SYSTEM SCALES		YOUR SCORE	MAX SCORE	
EMERGENCY				- +
()	1 CARDIOVASCULAR	46	56	***S***...
(+)	2 RESPIRATORY	33	36	*****S...
()	3 GASTROINTESTINAL	27	32	*****S**...
(-)	4 NEUROLOGY	24	32	S***.....
ACUTE				
(+)	5 CARDIOVASCULAR	66	68	*****S.
()	6 MUSCULOSKELETAL	46	56	S***...
(-)	7 RESPIRATORY	52	64	S***...
()	8 GASTROINTESTINAL	56	64	* **S**...
(-)	9 GENITOURINARY	39	52	S*****...
()	10 NEUROLOGY	42	48	***S*...
()	11 ENDOCRINE	14	16	***S*...
(-)	12 DERMATOLOGY	36	48	S*****...
()	13 EYES & ENT	38	48	S*****...
()	14 HEMATOLOGY	24	32	*S***.....
()	15 REPRODUCTIVE	44	48	*S....
()	16 PSYCHOSOCIAL	45	52	**S***...
CHRONIC				
()	17 CARDIOVASCULAR	42	48	***S*...
()	18 MUSCULOSKELETAL	33	40	**S***...
(-)	19 RESPIRATORY	28	40	S...****...
()	20 GASTROINTESTINAL	38	48	S*****...
()	21 GENITOURINARY	25	32	**S**.....
()	22 NEUROLOGY	33	36	*****S...
(-)	23 ENDOCRINE	43	56	S*****...
()	24 DERMATOLOGY	16	24	*S*****...
(-)	25 EYES & ENT	31	44	S*****...
()	26 HEMATOLOGY	22	28	**S*****...
(-)	27 REPRODUCTIVE	16	20	S*.....
(-)	28 PSYCHOSOCIAL	36	48	S.*****...
(-)	29 OTHER-(ECON,D IET)	44	64	S...*****...

QUALITY OF SYSTEM SCALE RESPONSES
PAGE 7 OF REPORT FOR AAPA ID NUMBER

IN TERMS OF SYSTEM SCALES, THE QUALITY OF YOUR RESPONSES APPEARS AS FOLLOWS:

H = HARMFUL U = UNACCEPTABLE

A = ACCEPTABLE E = EFFECTIVE

SYSTEM SCALES	MAXIMUM NUMBER OF RESPONSES	QUALITY OF RESPONSES			
		H	U	A	E
EMERGENCY					
1 CARDIOVASCULAR	14	1	3	1	.9
2 RESPIRATORY	9	0	1	1	7
3 GASTROINTESTINAL	8	0	2	1	5
4 NEUROLOGY	8	1	2	1	4
ACUTE					
5 CARDIOVASCULAR	17	0	1	0	16
6 MUSCULOSKELETAL	14	0	5	0	9
7 RESPIRATORY	16	1	4	1	10
8 GASTROINTESTINAL	16	0	4	0	12
9 GENITOURINARY	13	0	6	1	6
10 NEUROLOGY	12	0	3	0	9
11 ENDOCRINE	4	0	1	0	3
12 DERMATOLOGY	12	1	4	1	6
13 EYES & ENT	12	1	3	1	7
14 HEMATOLOGY	8	0	4	0	4
15 REPRODUCTIVE	12	0	2	0	10
16 PSYCHOSOCIAL	13	0	2	3	8
CHRONIC					
17 CARDIOVASCULAR	12	0	3	0	9
18 MUSCULOSKELETAL	10	1	2	0	7
19 RESPIRATORY	10	0	6	0	4
20 GASTROINTESTINAL	12	0	5	0	7
21 GENITOURINARY	8	0	3	1	4
22 NEUROLOGY	9	0	1	1	7
23 ENDOCRINE	14	0	6	1	7
24 DERMATOLOGY	6	0	2	0	3
25 EYES & ENT	11	0	6	1	4
26 HEMATOLOGY	7	0	3	0	4
27 REPRODUCTIVE	5	0	2	0	3
28 PSYCHOSOCIAL	12	0	6	0	6
29 OTHER-(ECON,DIET)	169	1	8	1	6

CONTINUED ON PAGE 8

SCORES RELATIVE TO ERROR PATTERN SCALE COMPETENCY BANDS

PAGE 8 OF REPORT FOR AAPA ID NUMBER

ERROR PATTERNS BASED UPON YOUR LESS THAN EFFECTIVE RESPONSES.
 IN TERMS OF ERROR PATTERN SCALES, YOUR ERROR SCORES IN RELATION
 TO A COMPETENCY BAND (*****) APPEAR BELOW.
 A MINUS SYMBOL BEFORE A SCALE NUMBER INDICATES THAT YOU MADE
 MORE THAN A REASONABLE NUMBER OF ERRORS.
 A PLUS SYMBOL BEFORE A SCALE NUMBER INDICATES THAT YOU MADE LESS
 THAN A REASONABLE NUMBER OF ERRORS.

ERROR PATTERN SCALES	YOUR ERROR SCORE	MAX ERROR SCORE	+
() 1 INSUFFICIENT KNOWLEDGE	8	60	..S*****.....
() 2 INSUFF DATA GATHERING	4	53	.S*****.....
(+) 3 INACCURATE DATA ANALYSIS	3	60	S*****.....
() 4 PREMATURE DIAGNOSIS	7	58	**S*****.....
() 5 INACCURATE DIAGNOSIS	7	60	**S*****.....
() 6 INCORRECT TREATMENT	10	60	****S**.....
() 7 INEFFECTIVE TREATMENT	4	61	S*****.....
() 8 DELAYED TREATMENT	5	46	..S*****.....
() 9 INAPPROPRIATE DRUG USE	8	63	**S***.....
(+) 10 INAPPROPRIATE REFERRAL	4	34	..S*****.....
() 11 INEFFECT INTERPERSONAL	6	31	..**S*****.....
(+) 12 INAPPROP INTERPERSONAL	1	38	S*****.....
() 13 EXCESSIVE DATA GATHERING	5	59	.S*****.....

END OF REPORT

III. NATIONAL PROFILES OF PROFESSIONAL COMPETENCE OF PHYSICIAN ASSISTANTS

1. Objectives

The development of national profiles of competence for a profession is a much bigger undertaking than is intended in this particular project. The purpose here is not to provide judgements on the competency strengths and weaknesses of the PA profession, but rather to identify areas of relative critical need for the development of 18 learning modules. This caveat is necessary in order to prevent the common error of over-interpreting data represented in such Reports. Professional competence entails a wide variety of knowledge, attitude and skills. The self-assessment examination by its very nature is limited in its approach in how much of this spectrum it can represent. These profiles should be considered relevant to the PA profession but they may by no means be comprehensive in representing it, nor absolute in judging it.

A profile has been described by Webster (1980, Second Edition) as a descriptive outline, in words or in graphic form, of the most outstanding characteristics of the subject of interest. The subject of interest is the professional competence of physician assistants in the United States. Given our limited interest in this instance, namely the

identification of deficiency areas in order to develop learning modules, a somewhat specific view of the subject is indicated. It was decided that a properly constructed paper-and-pencil self-assessment exam would provide sufficiently sensitive information for this purpose.

The process of profile generation entails

- a) the identification of the significant characteristics of the profession
- b) the development of an instrument that will be sufficiently sensitive in assessing these characteristics
- c) the administration of this instrument to an adequate national sample of the profession
- d) the verification of the representativeness of the data gathered (in terms of sampling bias and error),
- e) the development of verbal and visual outlines/summaries in order to represent the profession in terms of the selected characteristics, and
- f) the presentation of some guidelines by which the profile may be correctly interpreted and utilized.

Since Steps a and b, the identification of significant characteristics and development of the self-assessment examination, have been described elsewhere in the Report, this description will be limited to the following objectives:

- a) Administration of instrument to an adequate national sample
- b) Verification of representativeness (extent of sampling bias and error)
- c) Visual and Verbal outlines/summaries
- d) Application and Interpretation guidelines.

2. Methodology: Administration to National Sample

Adequacy of national sample is a matter which requires the consideration of sampling procedures and sample size. Stratification and randomness are usually desirable considerations, but often times there are other factors to consider such as, cost and convenience. The AAPA has developed a national data bank on its membership but it depends on a computer service company for its access to its own data bank. This process is costly and inconvenient especially when a previously undefined service is requested. Then there is the problem of randomness and representativeness. Currently the bank has 8000 names. However, it is believed that there are over 13,000 persons in the United States who could be labelled as 'physician assistants'. Not all PAs are members of AAPA, nor do all PAs get recorded in the AAPA data bank. Furthermore, the problem of keeping addresses current is also a very real one.

Because of these considerations, a systematic sampling procedure was requested. Since the computer service

company was accustomed to generating mailing labels it was decided to request a label for every other PA in the data bank, thereby producing an initial sample of about 4000 names. After screening these names for bad addresses and for persons who should not be included because they had taken the self-assessment exam previously or because they had been involved in its development, a list of some 3000 names was prepared.

HRA contract specifications indicated a sample size of 1500. In order to attain this number of usable exam answer documents, this list of 3000 names appeared appropriate. However, cost considerations required that the number who received exam booklets (cost of mailing exam booklet, answer sheets and return envelope was estimated at around \$2) be kept as close to the 1500 as possible. Accordingly a letter of inquiry was designed and sent to the 3000, under the AAPA President's signature, asking if the PA would be willing to participate in the self-assessment exam under the required deadlines, and, if so, to return a reply postcard and indicate address changes if any. The sample size thus dropped to 2200. Furthermore it now became classifiable as a "volunteer" group with all the limitations that this entails. Yet one could legitimately argue that all surveys are based on "volunteers"; those who do not volunteer just do not return their responses! In the case of this particular study, usable answer sheets were then received from about 1100 persons, which is exactly 50 percent of the sample that

was sent the exam materials, or about 37 percent of the selected sample.

Classical random selection procedure has of necessity been violated in the sampling method employed in this study. The adequacy of this approach is based upon its practicality, cost-effectiveness, and the fact that biases were avoided. Criteria for including persons in the sample were those minimally necessary for implementing the study. As far as possible, all sample selection judgements were based on AAPA ID Numbers and were made by clerical or technical persons uninvolved in AAPA politics.

How adequate is the sample size, namely 1500, prescribed by HRA's Contract guidelines? This is a difficult question to answer because it depends upon several factors, such as the nature of the variable that is of interest, its sampling distribution, and the degree of accuracy that is desired. In this study, our interest lies in the scale scores earned by PAs and in the proportion (p) of PAs who achieve minimum competency as defined for the entry-level generalist PA. If we assume that the proportion declared incompetent, say ' q ', is somewhat small, then the sampling distribution of ' q ' will be positively skewed and will approximate a Poisson distribution. Very large samples are needed in order to make decisions based on Poisson distributions and hence the justification for requiring a sample size of 1500.

3. Verification of Representativeness of Sample

It was recognized that the profiling of the PA profession would be based upon a sample that may not be entirely random or representative. Checks were therefore planned to verify this by including survey questions that would allow comparisons between the study sample and the PA profession survey conducted by the Association of Physician Assistant Programs (APAP) in 1978 (Perry and Redmond, 1980). The following types of data were gathered for comparison purposes:

- i) Breakdown of PAs by Sex
- ii) Year of graduation
- iii) Medical background prior to training
- iv) Location of practice
- v) Size of community where practice is located
- vi) Type of practice
- vii) Type of patient care provided
- viii) In-hospital location for patient care

The 1978 APAP National Survey (NS) was based on a sample of 4500 PAs who responded to it. There was no attempt to generalize these findings to the entire PA population in the United States which was estimated at over 8000 persons at that time. The focus of the study was on graduates of PA programs who constitute an increasingly major part of the PA profession.

The 1981 AAPA Self-Assessment Examination (SAE) PA

Profile (PAP) study is based upon 891 respondents out of 2200 who agreed to participate in the study and returned their responses in time for this analysis.

A comparison of responses to the eight types of PA characteristics, listed above, is presented in Tables 10-17. A chi-square test of significance was conducted to decide whether the SAE-PAP responses were similar in distribution to the APAP-NS responses. Outcomes of these tests are indicated as follows:

NS = Not Significant

* = Significant

Significance of the chi-square test rejects the hypothesis that the SAE-PAP responses are similar in distribution to the APAP-NS. Degrees of freedom for these tests depend on the number of response categories and have been indicated in each Table. A 5 percent level of significance was selected.

The results of the chi-square test of significance, indicate that the AAPA-SAE-PA National Profile sample of 891 is significantly different from the APAP - National Survey sample in terms of seven out of the eight PA characteristics that were selected. Only sex was non-significant, or similarly distributed in the two sample. The obvious conclusions to be drawn are:

- i) If the APAP - 1978 Sample is representative of the PA population, then the AAPA - SAE preliminary sample is not representative and needs to be ad-

justed appropriately. Such adjustments can be done statistically with the available data.

- ii) The APAP - 1978 Sample may be biased in favor of graduates from PA programs, in which case it would not be representative of the PA population. This means that the AAPA - SAE preliminary sample would need to be compared to another base line data-set, e.g. data from AAPA data bank, or data gathered by the National Commission for Certification of Physician Assistants.

Some major discrepancies in the characteristics of the two Samples are noted below:

- i) The SAE sample has a greater percentage of 1974-75 graduates, whereas the APAP sample has more 1976-77 graduates. This is understandable. The 1974-75 graduates are currently due to take the recertification exam (required every six years after certification) and would be more anxious to take the SAE. The APAP survey was conducted in 1978 and recent graduates are more likely to respond to requests for information by their training programs.
- ii) A larger percentage claimed 'other health fields' as their medical background prior to PA training in the SAE-PAP sample whereas the APAP-NS had more Medical Technicians and Corpsmen. This appears to be a definite shift occurring in the composition of the PA profession as it matures and comes into its own, a

TABLE 10
BREAKDOWN OF PAs BY SEX

<u>Response</u>	SAE - PAP 1981 (N = 884)	APAP - NS 1978 (N = 4469)
A. Male	70.0	69.4
B. Female	29.2	30.6
Chi-square = 2.99 NS (df = 1)		

TABLE 11
YEAR OF GRADUATION

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 3648)
A. Before 1972	3.0	0.1
B. 1972 or 1973	13.2	4.2
C. 1974 or 1975	36.0	23.5
D. 1976 or 1977	29.6	43.6
E. After 1977	17.8	25.4
Chi-square = 317.26 * (df = 4)		

TABLE 12
MEDICAL BACKGROUND PRIOR TO TRAINING

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 4500)
A. Medical Corpsman	36.7	41.9
B. Medical Technician or Technologist	13.4	20.1
C. RN	10.1	13.4
D. Other health field	24.6	13.2
E. Other non-health field	14.5	11.3

Chi-square = 128.95 *
(df = 4)

TABLE 13
LOCATION OF PRACTICE

<u>Response</u>	U.S. POPULATION (N = 203.2 Million)	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 3108)
A. Northeast	24.1	27.9	22.1
B. South	30.9	25.0	32.8
C. North Central	27.9	18.5	22.4
D. West	17.1	24.0	22.7

Chi-square = 69.07 *
(df = 3)

Chi-square = 36.86 *
(df = 3)

TABLE 14
SIZE OF COMMUNITY WHERE PRACTICE IS LOCATED

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 3320)
A. Under 10,000 persons	27.7	24.6
B. 10,001 to 50,000	26.9	21.5
C. 50,001 to 250,000	19.4	17.9
D. Over 250,000	23.7	36.0

Chi-square = 55.75 *
(df = 3)

TABLE 15
TYPE OF PRACTICE

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 3618)
A. Solo practice	18.7	17.9
B. Group practice	20.1	16.8
C. Hospital	26.9	23.7
D. Non-hospital clinic setting	21.9	23.4
E. Military setting	10.3	18.2

Chi-square = 41.63 *
(df = 4)

TABLE 16

TYPE OF PATIENT CARE

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 3416)
A. Family/General Medicine, including Ob/Gyn	48.3	53.9
B. Surgery	11.0	11.7
C. Pediatrics	4.9	3.3
D. Emergency Medicine	10.9	4.9
E. General Internal Medicine	22.1	12.0
Other Medical Specialties	2.8	6.3
Other specialties		7.8

Chi-square = 82.31 *
(df = 4)

TABLE 17

IN-HOSPITAL LOCATION FOR PATIENT CARE

<u>Response</u>	SAE - PAP 1981 (N = 891)	APAP - NS 1978 (N = 2611)
A. Inpatient bedside	28.7	35.0
B. Emergency room	12.5	25.4
C. Outpatient clinic	34.0	22.0
D. Operating room	4.9	10.2
E. Other areas	12.6	7.4

Chi-square = 192.57 *
(df = 4)

trend that was predicted by the APAP-NS study (Perry and Redmond, 1980).

- iii) The APAP-NS Survey had a greater preponderance of large-city practitioners, whereas the SAE-PAP group is more evenly distributed with a slight edge towards the smaller-town/rural practitioners. The APAP-NS did note an increasing trend for PAs to be hired by over-worked family physicians in smaller communities (Ibid, p.81). But the study also indicated a modest movement towards urban areas. It is possible that practice location is being influenced more by job opportunity than by personal PA preference.
- iv) The geographic regions breakdown showed fewer persons from the South and North Central regions for the SAE-PAP group.
- v) The SAE-PAP group had proportionately more PAs practicing Emergency Medicine than the APAP-NS group. Otherwise the two groups were quite comparable. However, the trend towards Emergency Medicine was anticipated by the APAP-NS study (p.81).
- vi) There were fewer PAs in military settings and more PAs in group practice settings in the SAE-PAP sample. This appears to be a trend that was recognized in the APAP-NS study.
- vii) The SAE-PAP group had a much larger proportion practicing in out-patient clinics, again a trend that

appears believable. Very few were from Operating Room type of hospital-based practice. This finding is somewhat difficult to integrate, given the Perry and Redmond study prognosis (p.85) that increasing numbers of PAs would be employed in surgery.

4. Visual and Verbal Summaries

In order to provide a competency profile of the PA profession it was decided to gather data pertaining to the following types of issues:

- a. Overall extent of competency attainment
- b. Specific strengths and weaknesses
- c. Relationship of practice characteristics to competencies profile
- d. Relationship of perceived educational needs to competencies profile

Competency was generally defined in terms of performance on the self-assessment examination. More specifically, it was operationalized in terms of scores on the 17 Role Scales, the 28 System Scales, and the 13 Error Pattern Scales.

Competency Standards were developed at two levels, an Entry level and an Advanced level, using a modified Nedelsky Technique. The Entry Level standards were judged to be relevant to the competency profile development process. For the purposes of this study, Entry-Level is to be interpreted

as the expected level of competence for a generalist PA not only at the time of entry into practice but at all times thereafter. The Advanced-Level is a tentative indicator developed on an experimental basis in order to reflect the level of PAs with significant experience in certain specialized practice areas.

Entry level standards for the 17 Role Scales, along with reliability indicators, are presented in Table 18. Tables 19 and 20 present similar data for the 28 System Scales and the 13 Error Pattern Scales. The entry-level standard provides the cut-off by which competency in the specific scale is judged.

Overall Extent of Competency Attainment

It is important to ask whether the PA profession has attained competency on an overall basis in the self-assessment exam. We can answer this question by examining the number of scales (out of 17 in the case of the Role Scales) that a PA passed or attained competency in. Figures 7 and 8 present the number of scales passed from the Role Scales, and System Scales. Ideally, all PAs should be expected to pass all the scales of the exam. In practice, recognizing that professionals tend to concentrate their practice in specific areas, it is recognized that certain areas might weaken through disuse. This is not excusable, but is recognized as something that happens and is therefore to be recognized when new learning modules have to be developed.

TABLE 18

ENTRY LEVEL STANDARD AND RELIABILITY INDEX FOR ROLE SCALES

Codes Used: r_{kr20} = Kuder-Richardson reliability index
 CA = Coefficient of Acceptance (a more appropriate
 reliability index for competency tests)
 SEM = Standard error of measurement

SCALE #	# ITEMS	MAX SCORE	ENTRY LEVEL STANDARD	r_{kr20}	CA	SEM
1	5	20	17.00	15	91	1.08
2	14	56	45.33	23	92	1.65
3	18	72	58.83	42	73	2.03
4	15	60	45.17	21	95	1.79
5	29	116	98.33	33	59	2.68
6	18	72	59.92	30	61	1.99
7	30	120	94.83	50	67	2.75
8	13	52	40.33	31	75	1.91
9	23	92	73.50	31	69	2.35
10	17	68	54.33	30	52	1.98
11	23	92	76.33	43	62	2.41
12	18	72	55.17	33	80	2.21
13	30	120	94.75	37	69	2.64
14	16	64	52.75	22	78	1.87
15	5	20	18.33	28	77	0.99
16	10	40	32.75	8	79	1.42
17	20	80	62.33	26	57	2.18

TABLE 19

ENTRY LEVEL STANDARD AND RELIABILITY INDEX FOR SYSTEM SCALES

Codes Used: r_{kr20} = Kuder-Richardson reliability index
 CA = Coefficient of Acceptance (a more appropriate
 reliability index for competency tests)
 SEM = Standard error of measurement

SCALE #	# ITEMS	MAX SCORE	ENTRY LEVEL STANDARD	r_{kr20}	CA	SEM
1	14	56	41.08	21	90	1.82
2	9	36	26.92	22	92	1.45
3	8	32	23.50	30	94	1.43
4	8	32	25.83	15	62	1.22
5	17	68	53.00	41	92	2.08
6	14	56	47.83	30	66	1.84
7	16	64	55.00	45	56	2.08
8	16	64	52.67	17	75	1.91
9	13	52	41.67	8	53	1.42
10	12	48	37.67	28	90	1.70
11	4	16	13.00	7	85	0.75
12	12	48	38.00	18	60	1.69
13	12	48	38.83	34	77	1.65
14	8	32	24.17	30	89	1.39
15	12	48	43.83	23	56	1.61
16	13	52	43.00	9	71	1.49
17	12	48	38.33	28	73	1.66
18	10	40	32.00	42	78	1.44
19	10	40	35.17	7	50	1.30
20	12	48	39.08	26	77	1.65
21	8	32	24.17	19	84	1.30
22	9	36	25.92	10	72	1.33
23	14	56	47.83	28	62	1.85
24	6	24	17.17	22	95	1.24
25	11	44	33.33	18	59	1.22
26	7	28	21.33	19	67	1.35
27	5	20	17.50	22	60	1.10
28	12	48	40.33	20	68	1.64
29	16	64	52.50	20	71	1.89

TABLE 20

ENTRY LEVEL STANDARD FOR ERROR PATTERN SCALES

SCALE #	# ITEMS	MAX SCORE	ENTRY LEVEL STANDARD
1	60	60	20
2	53	53	17
3	60	60	18
4	58	58	17
5	60	60	18
6	60	60	13
7	61	61	20
8	46	46	18
9	63	63	11
10	34	34	15
11	31	31	12
12	38	38	7
13	59	59	16

FIGURE 7

NUMBER OF ROLE SCALES PASSED AT COMPETENCY LEVEL BY SAE-PAP SAMPLE

NUMBER OF ROLE SCALES PASSED	PERCENT OF PAS IN SAMPLE (N = 891)		
	10	20	30
1	.		
2	.		
3	.		
4	.		
5	.	.	
6	.	.	
7	.	.	
8	.	.	.
9	.	.	.
10	.	.	.
11	.	.	.
12	.	.	.
13	.	.	.
14	.	.	.
15	.	.	.
16	.	.	.
17	.	.	.

FIGURE 8

NUMBER OF SYSTEM SCALES PASSED AT COMPETENCY LEVEL BY SAE-PAP SAMPLE

NUMBER OF SYSTEM SCALES PASSED	PERCENT OF PAS IN SAMPLE (N = 891)		
	10	20	30
1 .			
2 .			
3 .			
4 .			
5 .			
6 .			
7 .			
8 .			
9 .			
10 .			
11 .			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26 .			
27 .			
28 .			

For the purposes of this study, Figures 7 and 8 reassuringly present negatively skewed distributions. They also indicate that the typical PA is competent in at least 13 out of the 17 Role Scales (76 percent) and 21 out of the 28 System Scales (75 percent). In comparing these findings with those from the pilot study (D'Costa and Fisher, 1980) it is noted that the pilot study found similar although somewhat more optimistic results. This could be because of the sample in the pilot study or because the standards set were less rigorous than those set for Form A4.

Specific Strengths and Weaknesses

Figures 9, 10 and 11 present frequency distributions indicating the percent of PAs who attained the entry level competency standard on each of the three types of Scales -- Role, System, Error Patterns. Using the same 75 - 80 percent expectation for the profession as in the pilot study, Role Scales for which CME programs need to be developed are identified. In rank order of critical need for learning module development are the following four scales:

10. Analyze data - Psycho-social, Neurology
5. Gather data - Psycho-social, Neurology
6. Analyze data - Endocrine
11. Formulate plan - Musculo-Skeletal

FIGURE 9 PERCENT OF PAs WHO 'PASSED' EACH ROLE SCALE AT ENTRY-LEVEL COMPETENCY

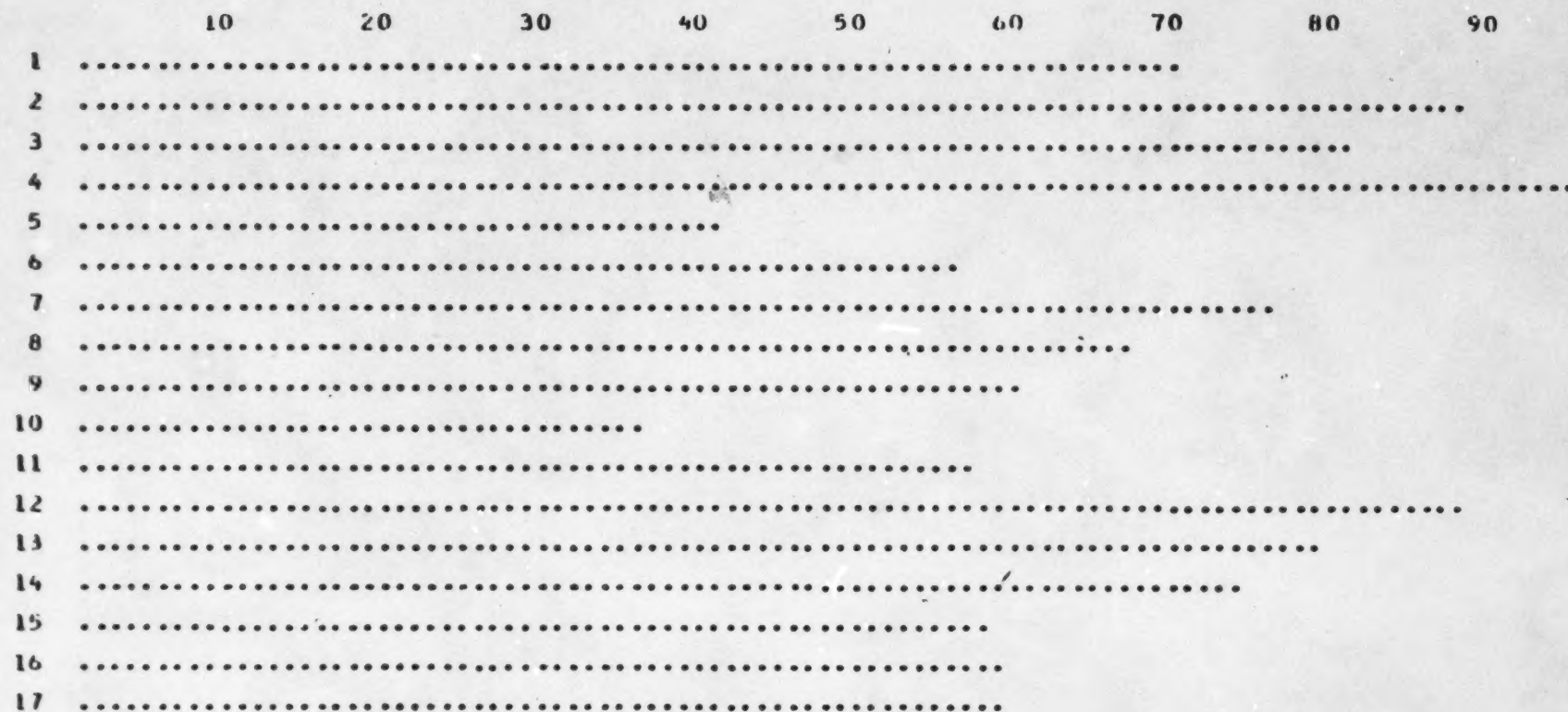


FIGURE 10 PERCENT OF PAS WHO 'PASSED' EACH SYSTEM SCALE AT ENTRY-LEVEL COMPETENCY

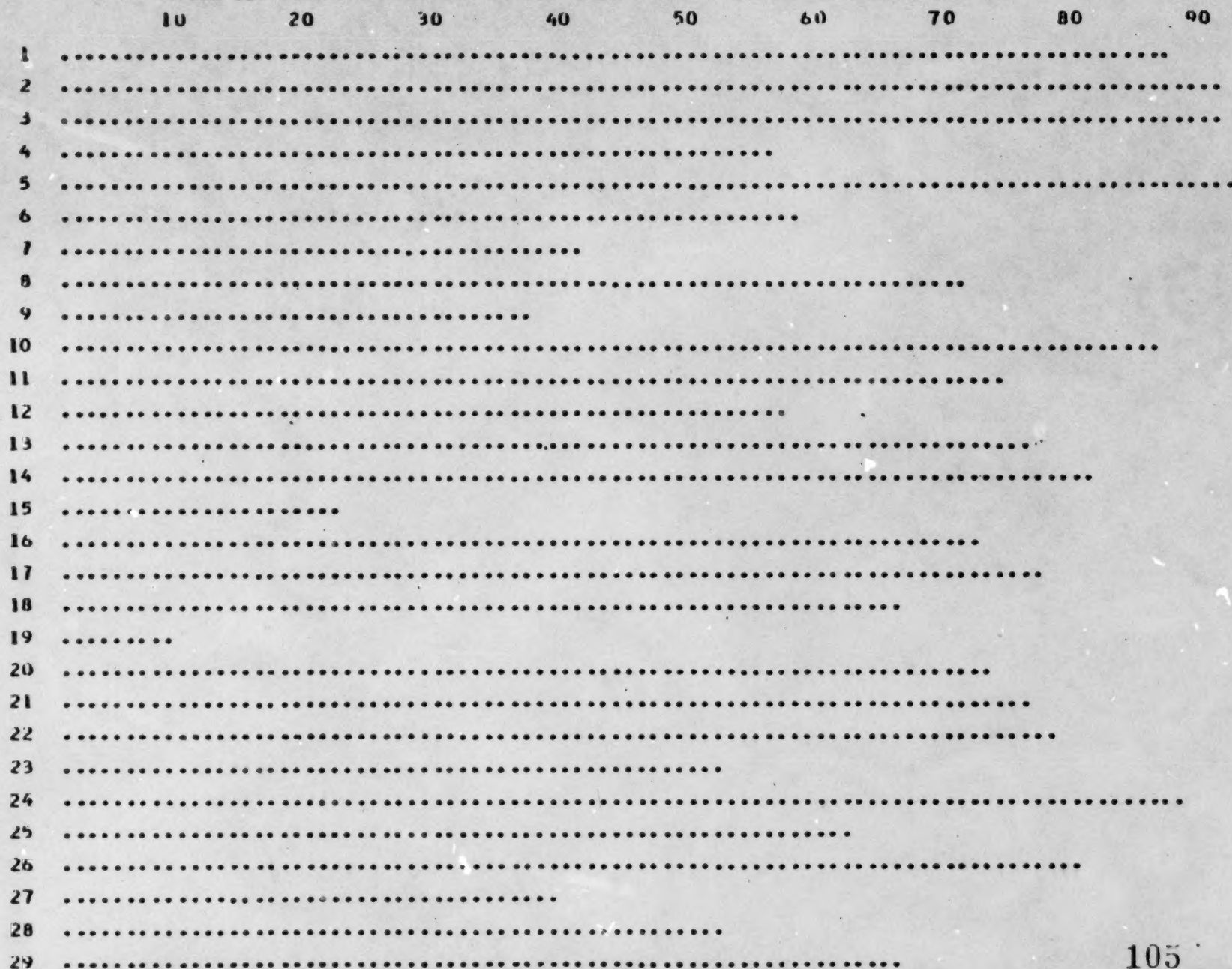
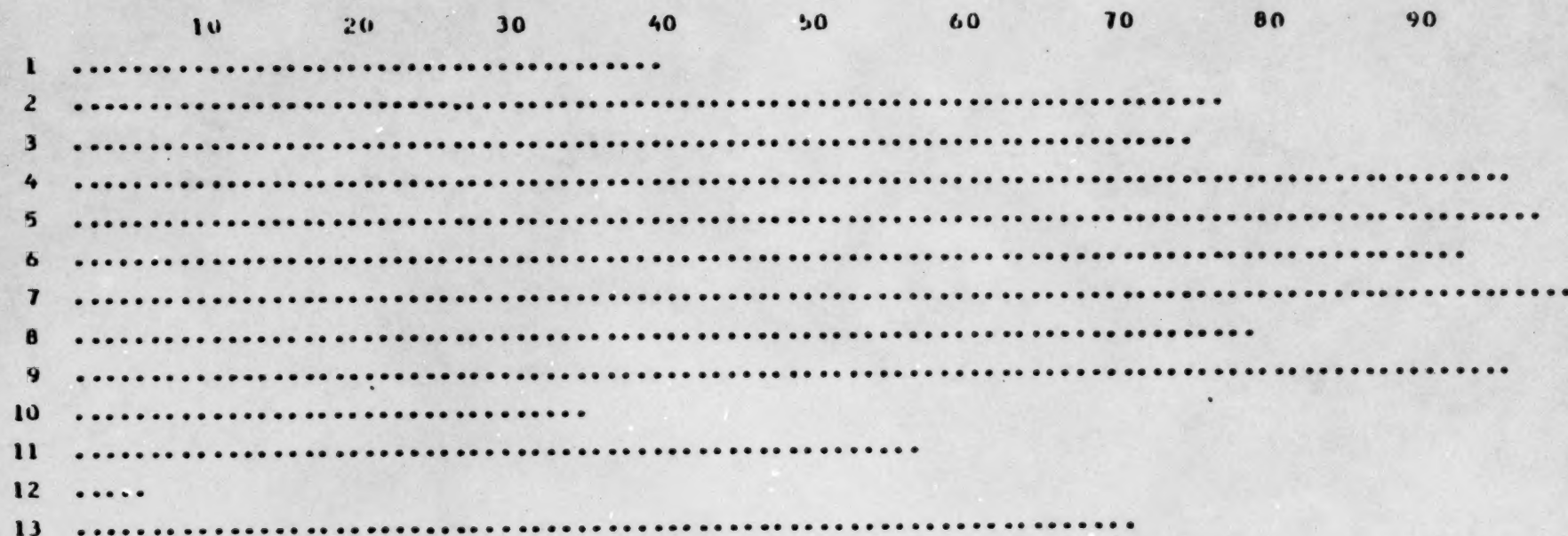


FIGURE 11 PERCENT OF PAs WHO 'EXCEEDED' ADVANCED-LEVEL ERROR PATTERN STANDARD

(Note: Exceeding standard implies excessive errors or a weakness)



A similar priority list for the System Scales would be:

- 19. Chronic - Respiratory
- 15. Acute - Reproductive
- 27. Chronic - Reproductive
- 9. Acute - Genito-Urinary
- 7. Acute - Respiratory

Relationship of Practice Characteristics to Competencies Profile

The purpose of this analysis is to provide a brief overview of the relevance of the observed competency deficiencies to PA professional practice needs. If the deficiencies occur in the same areas that the PA's practice is characterized by, then there is greater significance to the development of learning modules in these areas. Table 21 presents the list of practice characteristics in rank order. It is noted that the five most significant practice areas are:

- i) Chronic Cardiovascular
- ii) Acute Musculoskeletal
- iii) Chronic Musculoskeletal
- iv) Chronic Respiratory
- v) Chronic Endocrine

Relationship of Perceived Educational Needs to Competencies Profile

In this analysis it is assumed that perceived edu-

cational needs expressed by PAs should be taken into account in developing learning modules. Table 22 presents the list of expressed needs in rank order. The five most significant needs areas are:

- i) Acute Hematology
- ii) Acute Endocrine
- iii) Emergency Cardiovascular
- iv) Emergency Neurology
- v) Emergency Respiratory

5. Application and Interpretation Guidelines

Criteria for Learning Module Identification

All paradigms for instructional development include a step which requires that the learner be studied. Yet most examples of actual instructional development efforts show few data about learners. This project, the Pa Self-Assessment Project, is in striking contrast. Data are available from the administration of the self-assessment instrument. These data were used to suggest instructional module development. Table 23 summarizes the need for continuing education modules by PAs with respect to the 29 System Scales by using five criteria:

- i) Practice (identifies most common practice types)
- ii) Perceived Educational Need (identifies areas of greatest discomfort)

TABLE 21

LIST OF PRACTICE CHARACTERISTICS IN RANK ORDER

MOST PATIENTS		'MEAN' NUMBER OF PATIENTS SEEN
	CHRONIC CARDIOVASCULAR	11-40
	ACUTE MUSCULOSKELETAL	11-15
	CHRONIC MUSCULOSKELETAL	9-12
	CHRONIC RESPIRATORY	9-12
	CHRONIC ENDOCRINE	9-12
	CHRONIC GENITOURINARY	9-12
	ACUTE GASTROINTESTINAL	7-10
	ACUTE GENITOURINARY	7-10
	ACUTE RESPIRATORY	7-10
	ACUTE CARDIOVASCULAR	10
	CHRONIC NEUROLOGY	4-8
	ACUTE EYES & ENT	5-7
	CHRONIC GASTROINTESTINAL	4-8
	CHRONIC EYES & ENT	5-7
	ACUTE DERMATOLOGY	5-7
	ACUTE NEUROLOGY	4-6
	CHRONIC PSYCHOSOCIAL	3-4
	CHRONIC DERMATOLOGY	3-4
	OTHER (ECON., DIET)	3-4
	ACUTE REPRODUCTIVE	3-4
	EMERGENCY RESPIRATORY	2-3
	EMERGENCY GASTROINTESTINAL	2-3
	EMERGENCY CARDIOVASCULAR	2-3
	EMERGENCY NEUROLOGY	2-3
	CHRONIC REPRODUCTIVE	2
	ACUTE PSYCHOSOCIAL	2
	CHRONIC HEMATOLOGY	2
	ACUTE ENDOCRINE	0-1
	ACUTE HEMATOLOGY	0-1
FEWEST PATIENTS		

TABLE 22

LIST OF PERCEIVED CARE NEEDS IN RANK ORDER

MOST UNCOMFORTABLE



ACUTE HEMATOLOGY
 ACUTE ENDOCRINE
 EMERGENCY CARDIOVASCULAR
 EMERGENCY NEUROLOGY
 EMERGENCY RESPIRATORY
 CHRONIC HEMATOLOGY
 ACUTE CARDIOVASCULAR
 ACUTE PSYCHOSOCIAL
 CHRONIC REPRODUCTIVE
 ACUTE NEUROLOGY
 CHRONIC GASTROINTESTINAL
 CHRONIC NEUROLOGY
 CHRONIC ENDOCRINE
 CHRONIC PSYCHOSOCIAL
 OTHER (ECON., DIET)
 EMERGENCY GASTROINTESTINAL
 ACUTE RESPIRATORY
 ACUTE REPRODUCTIVE
 CHRONIC DERMATOLOGY
 CHRONIC RESPIRATORY
 ACUTE GASTROINTESTINAL
 CHRONIC CARDIOVASCULAR
 ACUTE EYES & ENT
 ACUTE DERMATOLOGY
 CHRONIC GENITOURINARY
 ACUTE GENITOURINARY
 CHRONIC EYES & ENT
 CHRONIC MUSCULOSKELETAL
 ACUTE MUSCULOSKELETAL

LEAST UNCOMFORTABLE

TABLE 23

MODULE DATA ANALYSIS

SYSTEM SCALE	PRACTICE PROFILE X = COMMON	NEED PROFILE X = UNCOMFORTABLE	ROLE SCALE NEEDS X = < 70% PASSED XX = < 50% PASSED	SYSTEM SCALE NEEDS X = < 70% PASSED XX = < 50% PASSED	NUMBER QUESTIONS TEST	PRIORITY NEEDS	MODULE PROPOSED
<u>EMERGENCY</u>							
1 Cardiovascular					12		
2 Respiratory					10		
3 Gastrointestinal			X		8		
4 Neurology		X	XX	X	8	3	
<u>ACUTE</u>							
5 Cardiovascular		X			16		
6 Musculoskeletal	X		X	X	14	1	3
7 Respiratory	X	X		XX	16	1	4
8 Gastrointestinal	X		X		16	3	
9 Genitourinary	X		X	XX	12	2	1
10 Neurology		X	XX		12	2	1
11 Endocrine		X	X		10		
12 Dermatology	X			X	12	3	
13 Eyes & ENT	X				12		
14 Hematology		X	X		8		
15 Reproductive	0-1 or >7		X	XX	12	2	1
16 Psychosocial			XX			3	
112							113

TABLE 23 (Cont'd.)

MODULE DATA ANALYSIS

99

SYSTEM SCALE	PRACTICE PROFILE X = COMMON	NEED PROFILE X = UNCOMFORTABLE	ROLE SCALE NEEDS X = < 70% PASSED XX = < 50% PASSED	SYSTEM SCALE NEEDS X = < 70% PASSED XX = < 50% PASSED	NUMBER QUESTIONS TEST	PRIORITY NEEDS	MODULE PROPOSED
CHRONIC							
17 Cardiovascular					12		
18 Musculoskeletal	X		X	X	10	2	1
19 Respiratory	X		X	XX	10	1	2
20 Gastrointestinal		X	X		12	3	
21 Genitourinary			X		8		
22 Neurology		X	XX		9	3	
23 Endocrine	X	X	X	X	9	1	2
24 Dermatology	X				6		
25 Eyes & ENT	X			X	11	2	1
26 Hematology		X	X		7		
27 Reproductive		X		XX	5	3	
28 Psychosocial	X	X	XX	X	12	1	2
29 Other	X	X	X		16		
114							115

- iii) Role Scales Competency (identifies deficiencies noted)
- iv) System Scales Competency (identifies deficiencies noted)
- v) Number of Exam Items per System Scale (provides an index of confidence in findings)

Based upon these five criteria, a three-level system of Priority Needs was identified and applied to the obtained data. The outcomes of this step are shown in Column 7 of Table 23. Module development was prioritized by project staff by identifying Body System Scales showing need or deficiencies in several of the criteria. Then the 18 contracted-for modules were apportioned across the first and second priority systems.

Interpretation Guidelines

One major concern relates to the entry-level competency standards that have been the bases for the national profiles. The "standards" utilized in this study were based on the individual judgements of the Advisory Committee, three of whom are PAs. The question as to whether these standards are appropriate for the total profession is open to debate. One criticism is the elite quality of the PAs in the Committee. On the other hand, one can reasonably argue that standards for PAs must be set by

physicians since they hire and supervise PAs. Also, one could reasonably argue that it is the elite segment of any profession that carries forward its torch and defends its prestige.

If the information and data presented in this section is interpreted as reflecting the status of the PA profession, care is certainly needed. The limitations must be recognized so as to preclude judgements of the PA profession. As a mechanism for identifying learning modules, it is believed that this Report is adequate for the following reasons:

- i) Strictness of standards will largely affect the length of the list of areas that recommend themselves for learning module development. One could choose the most pressing needs and bypass the problem of stricter-than-needed standards.
- ii) Bias in standards would be a different type of problem. This could arise because of specialties represented in the Advisory Committee. It is believed that the composition of the Advisory Committee precluded such bias.
- iii) An inadequate sample would affect the reliability of the performance findings derived from the profile data. It is believed that the obtained sample of 891 is certainly not small. Furthermore comparisons of results based on this sample of 891 with those based on previous pilot study samples of 108

and 210 suggest that the findings are indeed comparable and stable.

- iv) A biased sample would affect the validity of the findings of this study. This does not appear to be a serious problem at this time although this profile sample does appear significantly different from the APAP - 1978 National Survey sample. The differences do appear to be reasonable to expect at this point in time. Given the fact that the PA profession is in a state of flux because of its youthfulness the obtained sample may indeed represent its current status. However, further data will be gathered to shed light on this matter.

IV. DEVELOPMENT OF LEARNING MODULES

1. Objectives

The goal of the self-assessment process is to identify areas of strengths and weaknesses for Physician Assistants, so that they may study their deficient content areas and as a result, upgrade the quality of the care they provide. An important step in the continuing education process becomes prescribing sources of remediation for PAs who have been identified as deficient in a certain area(s). Towards this goal, the present project was contracted to develop 18 self-paced learning modules, covering areas identified by looking at the weaknesses of a sample of the PA profession as measured by the self-assessment exam.

The modules were designed to fit easily into a PA's schedule. They are self-paced and are designed to enable the individual to fit learning into individual constraints on time and use of space. The module format also allows PAs to pick and choose content areas that are essential for their particular professional situation.

The style of the modules is varied depending on the nature of the subject and the style of the authors. The instructional strategies range from that of programmed instruction through Ausubel-inspired advance organizers, and Keller-style directions and quizzes to linear monographs.

This variety should ensure that a number of learner cognitive styles are accommodated.

2. Methodology

The modules have been developed using systematic methods as outlined below.

Identification and Prioritization of Areas of Deficiency

Because the areas to be used as topics for modules were to be those in which PAs demonstrated a need or deficiency, the choice of module topics depended on data from a sample of PAs taking the self-assessment exam. The SAE was sent to 2200 PAs in January 1981. 888 PAs completed and returned the exam by March 31, 1981. This group formed the sample used for purposes of compiling a descriptive profile of the profession. Admittedly, this is not a true random sample in that a bias in terms of who completed the exam may have been operating, but the representativeness of the sample has been checked and presented earlier in this Report.

Information from this sample was tabulated in a chart format and utilized to prioritize areas of need by Body System topic areas. In particular, the following information was noted. System scale areas were identified with an X that the sample had responded to as those in which they

saw the most patients. System areas were also identified with an X that the sample had responded to as those in which they were the most uncomfortable. Additionally, system and role scale areas were identified as those in which <70% (XX) and <50% (X) passed the competency standard set for an entry level PA. The number of items in the test in each scale area were also recorded. These five columns of Xs (# patients seen in an area, degree of uncomfortableness in an area, % passing the standards, and number of items in the scale area) were considered by the staff, and the system areas were prioritized in terms of need for module coverage as 0,1,2,or 3. The staff used these priority ratings as a basis for proposing a certain number of modules that should be developed for each system area. The total number of modules proposed was 18. This chart is reproduced in Table 23.

These proposed areas for module development were presented to the Advisory Committee at the March 1981 meeting. The Advisory Committee agreed that modules were needed in the system areas proposed by the staff. They produced a list of content experts who would be appropriate as module authors. These experts were contacted and eight agreed to participate in the training workshop. Two Advisory Committee members also agreed to participate as potential authors.

Training of Authors

The group of authors was brought together at the Fawcett Center for Tomorrow, the OSU Continuing Education Center, for a two day workshop in April, 1981 on the design of self-paced learning modules. At this time Dr. John Belland (Project Consultant) and the project staff discussed the following topics:

- 1) Data on the percent of PAs who had passed the competency standard for each of the 29 Body System areas, and the items and item statistics on each system scale targeted for module development
- 2) The nature of the instructional process, and style requirements for the modules
- 3) Assignment of authors to topics, author responsibilities and the stages of development the modules were to go through
- 4) Content expert resources for each module
- 5) Plans for the writing workshop

A pertinent set of readings was mailed to the authors before the workshop, and a Manual of Procedure and Style was provided to each author at the workshop. Authors were given 30-60 days to gather references and organize their thinking before meeting again at the writing workshop.

Learning Modules Draft: Writing Workshop

The authors were brought back together at the Fawcett Center on June 8-11, 1981 for a four day working session during which the authors began their first draft of the modules with the assistance of the project staff and consultant. OSU staff provided the authors with: 1) typewriters, word-processing, and dictation/transcription equipment, 2) access to the OSU Health/Science library, 3) consultation with not only project staff but also OSU faculty and staff resources, and 4) consultations with a medical illustrator. Authors paced themselves and worked in a variety of locations, but were in contact with the project staff at least twice each day.

Some of the authors completed their first draft during the June workshop, and the others were scheduled to have their first drafts into the project office by July, 1981. However, a number of the authors did not make this deadline, which began to put the project behind its schedule. The staff worked with a medical illustrator to produce the illustrations desired by the authors. They obtained permission to use copyrighted material from publishers of figures indicated for inclusion in the modules by the authors.

As the authors finished the first draft, they were instructed to submit the modules to expert review and developmental testing. The purpose of these two processes was to improve the quality of the modules by obtaining an

outside check on the accuracy of the content (expert review) and on the readability and clarity of the presentation (developmental testing).

Expert Review

The authors were instructed to send their modules to at least one content expert who was not a party to the development of the module to assess the module for accuracy, completeness of content, and clarity of presentation. Some of the authors completed this task themselves and others asked the project staff to handle the expert review. The Advisory Committee met in September, 1981 and reviewed 8 of the modules. Their comments were relayed to the authors who made necessary revisions.

Developmental Testing

The authors were instructed to administer the module in person to one physician assistant who needed to learn the content of the module. The author was to observe the learner as he or she proceeded with the work of the module; noting any difficulties experienced by the learner, asking the learner to identify unclear portions of the modules, and noting these for future revisions. The authors were to make these revisions and then, iterate the process with other learners, with no additional reviews being required.

3. Results

The production of 18 Learning Modules was a huge undertaking requiring extensive communication with authors, with reviewers, with a medical illustrator, with publishing companies (to obtain copyright permission), and with the Advisory Committee. Twelve of the modules have been completed. These modules will be printed and made available to members of the AAPA. However, the remaining projected modules have been delayed due to inability of authors to complete work on schedule, and it is unknown as to the future of these modules. The twelve that are completed are presented below.

The 12 Self-Paced Learning Modules

1. Asthma, Judith B. Willis, M.A., PA-C, Wendell Weber, Ph.D., M.D., Geoffrey R. Grambau, M.D.
2. Perioperative Respiratory Care, J. Jeffrey Heinrich, PA-C, Shepard B. Stone, MPS, PA-C
3. Chronic Obstructive Pulmonary Disease, Geoffrey R. Grambau, M.D., John E. Schoell, M.D.
4. Data Gathering in Respiratory Disease, Geoffrey R. Grambau, M.D., John W. Dircks, M.D.
5. Preventive Screening for Common Visual Problems in Primary Care, Lawrence M. Kammer, M.S., PA-C, Michael P. McHenry, B.S., PA-C
6. Sexual Dysfunction, Ann E. Teske, R.N., Ph.D.
7. Abuse of Persons by Others, Ann E. Teske, R.N., Ph.D.
8. Low Back Pain Syndromes, C. Emil Fasser, PA-C

9. Drug Abuse, Wendell Weber, Ph.D., M.D.
10. Anxiety, C. Emil Fasser, PA-C, Wendell Weber, Ph.D., M.D.
11. Pulmonary Infection, Geoffrey R. Grambau, M.D., Thomas A. Abraham, M.D., Wendell Weber, Ph.D., M.D.
12. Cerebrovascular Diseases, Lawrence M. Kammer, M.S., PA-C, Michael P. McHenry, B.S., PA-C

V. DESIGN FOR CONTINUING COMPETENCY ASSURANCE PROGRAM

1. Objectives

The purpose of this design was to overlay on the existing system a reasonable and futuristic program of continuing competency assurance for PAs. Currently the PA profession is committed to a six year cycle of recertification. Every two years, each PA is required to log in 100 hours of approved CME credit. The AAPA monitors the quality of CME credits earned and provides the CME logging service. The recertification exam is offered by the National Commission for the Certification of Physician Assistants (NCCPA), an independent organization located in Atlanta, Georgia.

It was expected that the pressure to appear for a recertification exam every six years would reinforce the Academy's dedication to ensure continuing competence of its members, and that the self-assessment service would provide comprehensive competency assessment and learning prescriptions.

The objectives of the projected system were to be as follows:

1. It must permit individualized assessment that is relevant to the professional development of the PA.
2. It must be convenient to use, confidential, and non-threatening.

3. It must be based on sound educational principles so that motivation will occur and continuing competence will result.
4. It must utilize cost-effective and efficient computer technology.

2. Methodology and Results

The development of this system occurred as a result of discussions conducted among project Staff, Advisory Committee members, and AAPA representatives. It was decided that project resources would be utilized, within the contexts of the current contract, to move forward to developing a prototype system for self-assessment, practitioner feedback, and facilitation of CME. The following elements of the system have been tested and are partially in place:

a) A test item bank has been created. Currently the bank has three types of items: about 200 old items which need revision, 320 good items utilized in Form A4, and several thousand new items received from the University of Kansas item pool which require considerable work-up.

Item development is costly because it requires significant use of manpower. It was thought that CME credit might be provided to certain Academy members who are capable of and want to engage in revising the test items. This would provide a continuous stream of new and revised items at little or no cost to the Academy. Given the recertifica-

tion requirement, this option would be desirable to certain types of members, e.g. educators, researchers, specialists, who might find this a good learning vehicle.

Thoughts have also been given to a computer-based generation of test-items given the test "form" or "shell" provided by scenarios file. This is an innovative idea of the future which deserves to be pilot tested.

b) A self assessment system that is capable of individualized assessment must take into account the practice characteristics and professional needs of the PA. A beginning attempt has been made in this respect. The current self-assessment does indeed gather such data, but it does not adjust the exam on the basis of these practitioner characteristics or needs. Such a capability is possible to develop if the Academy moves to a more flexible modular exam. In requesting the self-assessment, either by mail or by computer terminal, the PA could indicate his/her background/needs and the system would return a special form of the exam. The system could be set up with a limited set of previously specified Exam Forms (derived from the Item Bank). With each request, a specific Form would be returned. Moreover, a Form could be randomly generated within the specifications already identified for the exam.

c) An educationally-sound system requires that it be based on sound theoretical principles and tested practices. The current system is based on a sound competency model derived from previous role delineation studies. It utilizes

tested diagnostic-prescriptive-learning practices. It requires learner-involvement and initiative. It is respectful of the individual PA's right to confidence and convenience. Psychological principles are used, such as when areas of strength were presented prior to areas of deficiency. Visual aids are emphasized so that the messages are clearly and quickly received.

All of these principles are already in place in the self-assessment system that has been pilot-tested. It is assumed that the individual will be motivated to use the system effectively. No follow-up or compliance strategies are currently utilized, although the Academy does log-in six hours of CME credit for the mere taking of the self-assessment exam. Thus the individual PA receives credit and visibility for using the service. However, there is no check to ensure that the results have been understood or that they have been meaningfully utilized.

It is hoped that 'policing' of this system will not become necessary in the future. Policing would destroy the voluntary nature of learning and defeat the principles on which the self assessment program is based. However meaningful "carrots" or reward systems could be built in, along with some 'sticks'. Both types of incentives are useful and professionals are no exception where need for motivation is concerned.

d) A periodic up-dating of the examination specifications and of the learning prescriptions is an essential

component of any dynamic system. Societal demands, practice characteristics, and technical expertise change. The role expectations of PAs are also likely to change. Already a trend towards specialty practice is emerging. Other practice trends could also emerge, e.g. emergency care, in which case competency demands would change. As PAs articulate with other health professionals, e.g., nurses, their role could get redefined, thereby leading to changes in the competency specifications.

e) The self-assessment system must be separate from the PA certification and training programs, yet there must be reasonable articulation among all three services. Self-assessment is clearly practitioner-oriented and the system must respond to practitioner needs. Training programs have their own mission, even though a part of it relates to relevance to later practice. Coaching students to take the self-assessment exam or the certification exam would defeat the broad purposes of PA training programs. Certification utilizes a very different philosophy of approach from that taken by a self-assessment exam. Complete truth-in-testing is possible, even emphasized in an exaggerated manner, in self-assessment. It would be horrendously expensive to implement similar practices in certification exams. The current self-assessment system permits the exam booklet to be retained by the individual PA. This means that new forms of the exam must be routinely created periodically. Such a policy is expensive, but essential to a self-assessment

service. Self assessment services are not threatened by exposure of exam forms. The competency model and the feedback strategy used in the self assessment system should be of interest to the other programs.

f) The self assessment system must utilize the latest computer technology available. Of special interest are home mini-computers which have now invaded many homes. Exams could be made available via diskettes (sent by mail), or via phone-connected computer terminals. Given the new technology, the exam could be made available upon individual demand in terms of specific modules. Upon completion of the exam, the individual could key-in the responses and receive immediate feedback with the 8-page Report being printed at his/her terminal.

VI. OVERALL RESULTS AND CONCLUSIONS

The AAPA-OSU-HHS partnership has combined the resources of a health profession, a state university, and the federal government to produce a viable model for a continuing competency assurance system for professionals. It is unfortunate that budget and time restrictions have constrained the development of a model system that has so far been well received. A unique approach to the development of self-assessment exams has been acquired which includes detailed feedback and learning prescriptions to the individual practitioner.

The system is inherently complex and therefore somewhat cumbersome to develop and maintain. The utility of the system is dependent on its proper utilization by individual practitioners, and the motivation for this must come both from within and outside the individual. Without some meaningful external "carrots and sticks" for recertification, professionals become victims of their own inertia and time-constraints. Interest in self assessment will increase as societal pressures for accountability take shape in professional realities. Self-assessment can be expected to become a meaningful approach to professional life span development. As micro-computers become available in every home, the ap-

approach piloted in this contract will appear even more viable, realistic, and useful.

Recommendations

The use of self-assessment approaches has long been recognized as an effective continuing competency assurance mechanism in the health profession. Using this system, a professional can identify strengths and weaknesses without risking humiliation and exposure, can be referred to appropriate remediation sources, and can review these resources at his or her own convenience.

The first and probably most important aspect of a self-assessment system is the adequacy of the assessment method. In the case of the PAs, the assessment method was operationalized in a criterion-referenced format. The criterion-referenced instrument has undergone changes and refinements as a result of this project, but there are a few questions still to be addressed.

One question that arises is whether the elaborate feedback (provided in terms of 17 Role Scales, 29 Body System Scales, and 13 Error Pattern Scales) is indeed useful to and utilized by the practitioner, and most importantly, whether a less elaborate feedback system might even be better, recognizing the problem of information overload. It has been suggested that a one-on-one process is needed to ensure that the practitioner will be motivated to pursue the

necessary follow-up.

A second question concerning the adequacy of the exam is in regard to the standards-setting issue. Because the PA's performance is classified as competent or non-competent according to where he or she scores relative to a standard defined as the score a minimally competent PA should obtain, it is important that these standards be as valid as possible. These standards have been set by a process followed by members of the Project Advisory Committee. As a result of the administration of the Form A4 developed during this project to a representative sample of PAs, data are available regarding the percent of PAs classified as competent and non-competent. These data could be a starting point for investigating the validity of the standards. The scales which indicate that a greater number of PAs were classified as non-competent, than PAs in the field would expect could be identified and the standards could be revised.

A third and important contribution made by this project was the production of a set of self-paced learning modules, covering content areas on which PAs as a group were weak as evidenced by the performance of a representative sample of PAs on Form A4 of the self-assessment exam. These learning modules provide short, concise reviews of content areas particularly relevant to PAs, and exhibit a variety of different formats (from programmed instruction to linear monographs).

The identification and development of sound learning

materials is an on-going process. The health professions need to utilize each others resources in this respect since modules can in certain cases be utilized across the professions. Greater use needs to be made of the AVLINE service provided by NIH so that such exchange becomes possible.

A fourth recommendation relates to the development of similar self-assessment systems for other health professions. This system is generalizable to other health professions. Perhaps the most important features to be emulated are the development of a competency model as a precursor to test development, the use of weighted scores and error patterns to increase the feedback provided to the practitioner, the use of the competency band approach, the comparative profile of practice, need, and exam scores and the providing of a detailed Individual Report with learning prescriptions. Each profession will need to adapt these ideas to suit their own goals and resources.

Finally, it needs to be said that the present AAPA continuing competency assurance program is far from the stage where it will stand on its own and continue its existence into the future. Additional forms of the exam are needed, and policy changes are needed both within AAPA and its related organizations in order to ensure the development of this system. The PA recertification program is still in its infancy and the ideal of continuing competency assurance is still tantalizingly elusive, as with most professions.

It is hoped that the gains made in this project will

be consolidated and that resources will be found to continue this effort forward.

References

- Berk, R.A. (Ed.) Criterion-Referenced Measurement. Baltimore, Md: The Johns Hopkins University Press, 1980.
- Carver, R.P. Special problems in measuring change with psychometric devices. In Evaluative research: Strategies and methods. Pittsburgh, Pa.: Amercian Institutes for Research, pp. 48-63.
- Conant, R.M. A credentialing concert. Journal of Environmental Health, 41 (2), 1978, 88-91.
- Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 16, 1951, 292-334.
- D'Costa, A. and Odor, P. A log-linear analysis of the relationships between practice needs, test-derived needs, and expressed needs for continuing medical education of Physician Assistants. A paper presented at the annual convention of AERA, New York City, April, 1982.
- D'Costa, A.G. and Watson, J.E. A critical incident technique for developing "criterion-referenced tests". A paper presented at the annual convention of AERA, Los Angeles, Ca., April, 1981.
- Flanagan, J.C. The critical incident technique. Psychological Bulletin, 51 (4), 1954, 327-359.
- Hambleton, R.K. Test score validity and standard-setting methods. In Criterion-Referenced Measurement, R.A. Berk (Ed.). Baltimore, Md.: The Johns Hopkins University Press, 1980.
- Hambleton, R.K. and Eignor, D.R. Criterion-Referenced Test Development and Validation Methods, AERA Training Program Materials, 1979.
- Huynh, H. On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, 13, 1976, 253-264.
- Messick, S.A. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 1975, 955-966.
- Nedelsky, L. Absolute grading standards for objective tests. Educational and Psychological Meeasurement, 14, 1954, 3-19.

Perry, H.B. and Redmond, E.L. Final Report. Deployment and Career Trends of Physician Assistants. Submitted to the Office of Health Research, Statistics, and Technology, National Center for Health Services Research, Department of Health and Human Services, 1980.

Popham, W.J. Criterion-Referenced Measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Popham, W.J. Domain specification strategies. In Criterion-Referenced Measurement, R.A. Berk (Ed.) Baltimore, Md.: The Johns Hopkins University Press, 1980.

Popham, W.J. and Husek, T.R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1969, 1-9.

Subkoviak, M.J. Estimating reliability from a single administration of a mastery test. Journal of Educational Measurement, 13, 1976, 265-276.

Swaminathan, H., Hambleton, R.K., and Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 11, 1974, 263-267.

The van de Kamp, Leo J. Generalizability and educational measurement. In Advances in Psychological and Educational Measurement, Dato N.M. De Gruijter (Ed.). London: John Wiley and Sons, 1976.

Wilson, M.A. Basic principles of credentialing health practitioners. Respiratory Care, 21 (10), 1976, 954.